15

N 67-15484 (ACCESSION NUMBER) 398 (PAGES)
(NASA CR OR TMX OR AD NUMBER)

N 67-15498 (THRU) 1 (CODE) 30 (CATEGORY)

FACILITY FORM 602

V.P. Vyazanitsyn and others

PHYSICS OF THE SOLAR SYSTEM

Volume 3 of

A Course in Astrophysics and Stellar Astronomy

Edited by A.A. Mikhailov

GPO PRICE $ _____

CFSTI PRICE(S) $ [redacted] 3.00

Hard copy (HC) _____

Microfiche (MF) $1 3.90

ff 653 July 65

Translated from Russian

Published for the National Aeronautics and Space Administration, U.S.A.

and the National Science Foundation, Washington, D.C.

by the Israel Program for Scientific Translations

V. P. VYAZANITSYN, M. N. GNEVYSHEV,
O. V. DOBROVOL'SKII, V. A. KRAT, A. V. MARKOV,
A. P. MOLCHANOV, V. M. SOBOLEV, and
V. V. SHARONOV

PHYSICS of the SOLAR SYSTEM

Volume 3 of

A COURSE IN ASTROPHYSICS AND STELLAR ASTRONOMY

(Kurs astrofiziki i zvezdnoi astronomii)

Edited by A. A. MIKHAILOV

Editorial Board: A. N. DEICH, V. A. KRAT,
O. A. MEL'NIKOV, and V. V. SOBOLEV

Izdatel'stvo "Nauka"
Moskva 1964

Translated from Russian

Israel Program for Scientific Translations
Jerusalem 1966

NASA TT F-342
TT 66-51022

Published Pursuant to an Agreement with
THE NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, U.S.A.
and
THE NATIONAL SCIENCE FOUNDATION, WASHINGTON, D.C.

Copyright © 1966
Israel Program for Scientific Translations Ltd.
IPST Cat. No. 1526

Translated by Z. Lerman
Edited by R. Hardin

Printed in Jerusalem by S. Monson
Binding: Wiener Bindery Ltd., Jerusalem

Price: $ 7.08

Available from the
U.S. DEPARTMENT OF COMMERCE
Clearinghouse for Federal Scientific and Technical Information
Springfield, Va. 22151

IX/13/3.5

CONTENTS

PREFACE 1

Part One
THE SUN

Chapter I. V. P. Vyazanitsyn. INTRODUCTION 3
§ 1. General data 3
§ 2. The total solar radiation 5
§ 3. Fluctuations in solar radiation 6
§ 4. The energy distribution in the solar spectrum 9
§ 5. Limb darkening 13
§ 6. Solar rotation 16

Chapter II. V. P. Vyazanitsyn. THE SOLAR LINE SPECTRUM 19
§ 7. The normal solar spectrum. Spectrum atlases and line catalogs 19
§ 8. The identification of Fraunhofer lines 22
§ 9. Equivalent Fraunhofer-line widths. Growth curves 25
§ 10. The quantitative composition of the solar atmosphere 32

Chapter III. V. A. Krat. THE STRUCTURE OF THE PHOTOSPHERE: GRANULATION, SUNSPOTS AND FACULAE 37
§ 11. The general appearance of the solar surface. Granulation 37
§ 12. Sunspots 43
§ 13. Solar rotation, as determined from spots and faculae 46
§ 14. Special features of the line spectra of spots and faculae 49
§ 15. Magnetic fields of spots. The general magnetic field of the sun 53
§ 16. Gas motion in sunspots. The Evershed effect 57
§ 17. The nature of sunspots. 59
§ 18. The cyclic variation of photospheric phenomena 61
§ 19. The cyclic nature of sunspot formation 66

Chapter IV. V. P. Vyazanitsyn and V. M. Sobolev. THE CHROMOSPHERE 68
§ 20. The flash spectrum. General properties of the chromosphere 68
§ 21. Spectrophotometric measurements and chromospheric-line profiles 72
§ 22. The chemical composition of the chromosphere 77
§ 23. The density distribution with height in the chromosphere 79
§ 24. Self-absorption in the chromosphere 84
§ 25. The temperature and the electron concentration. Chromospheric models 89
§ 26. Chromospheric spicules 98
§ 27. The ultraviolet radiation of the chromosphere 101
§ 28. The theory of chromospheric heating 103
§ 29. Bright flocculi 107

Chapter V. V. P. Vyazanitsyn and V. M. Sobolev. PROMINENCES 111
§ 30. Historical remarks 111
§ 31. Statistical data 112
§ 32. Filaments 114
§ 33. The classification of prominences 116
§ 34. The dynamics of prominences 123
§ 35. Prominence spectra. Spectrophotometric measurements 125
§ 36. Excitation and ionization 130

Chapter VI. V. A. Krat. CHROMOSPHERIC FLARES 134
§ 37. Classification. Frequency of flares 134
§ 38. The formation of surges 136
§ 39. The spectra of chromospheric flares and their interpretation 137
§ 40. Emission cores 141
§ 41. Geoactivity and the nature of flares 144

Chapter VII. V. A. Krat. THE SOLAR CORONA 148
§ 42. The continuous spectrum of the corona 148
§ 43. The line spectrum of the corona 151
§ 44. The brightness and polarization of the corona 153
§ 45. Structural features of the corona in "white" light 155
§ 46. The ionization and excitation of atoms in the corona 156
§ 47. Profiles of coronal lines. The structure of the corona in monochromatic emission of spectral lines 158
§ 48. Coronal motions. The development of coronal condensations 162
§ 49. The corona and the chromosphere 163

Chapter VIII. M. N. Gnevyshev. THE SOLAR SURVEY 169
§ 50. A definition of the term "Solar Survey" 169
§ 51. Photospheric solar-activity indexes 171
§ 52. Chromospheric solar-activity indexes 175
§ 53. Coronal solar-activity indexes 178
§ 54. Solar radio-emission indexes 179

Chapter IX. A. P. Molchanov. SOLAR RADIO EMISSION 182
§ 55. Methods of observation 182
§ 56. A general description of solar radio emission 185
§ 57. The undisturbed component of the solar radio emission 186
§ 58. The slowly varying disturbed component of the solar radio emission 208
§ 59. The rapidly varying disturbed component (radio bursts) 228

Part Two

THE PLANETARY SYSTEM

Chapter X. A. V. Markov. THE MOON 238
§ 60. The moon as a celestial body 238
§ 61. Photometric and polarization properties of the lunar surface 241
§ 62. Studies of the temperature of the lunar crust using heat receivers and radio methods . . 246
§ 63. Endogenous and exogenous hypotheses of lunar-relief formation 255
§ 64. Maps of the other side of the moon 259

Chapter XI. V. V. Sharonov. PHYSICS OF THE PLANETS 264
§ 65. Introduction 264
§ 66. The planetary disk 265
§ 67. Methods of observing planetary features 266
§ 68. Methods of disk measurement 267
§ 69. Phases 269
§ 70. Planetographic coordinates of surface points 270
§ 71. The rotation elements and their observational determination 272
§ 72. Spectroscopic studies of rotation 274
§ 73. The disk of a considerably flattened planet 277
§ 74. Planetary characteristics related to the mass 279
§ 75. The brightness, magnitude, and color of a planet 281
§ 76. Reflectivity 283
§ 77. Albedo 285
§ 78. The application of surface photometry to disks of planets and satellites 288
§ 79. Structures of planetary atmospheres 289
§ 80. Optical phenomena in planetary atmospheres 292
§ 81. Temperature conditions 294
§ 82. Planetary radio astronomy 296

Chapter XII. V. V. Sharonov. A DESCRIPTION OF INDIVIDUAL PLANETS 299
§ 83. Mercury 299
§ 84. Venus 300
§ 85. The earth 306
§ 86. Mars 308
§ 87. Jupiter 314
§ 88. Saturn and its rings 318
§ 89. Uranus, Neptune, and Pluto 321
§ 90. Planetary satellites 324

Chapter XIII. A. V. Markov. PHYSICAL PROPERTIES OF THE MINOR PLANETS 328
§ 91. Orbits of the minor planets 328
§ 92. Physical properties of the minor planets 329
§ 93. The origin of the minor planets 333

Chapter XIV. O. V. Dobrovol'skii. COMETS, METEORS, AND THE ZODIACAL LIGHT 336
§ 94. General data on comets 336
§ 95. Cometary nuclei 342
§ 96. Cometary spectra 347
§ 97. The apparent brightness of a comet. Masses and densities of cometary atmospheres . . 358
§ 98. Differentiation of matter in a cometary atmosphere under the influence of solar heat . . 361
§ 99. Type-I tails 363
§ 100. The origin of comets 369
§ 101. Some unsolved problems in the physics of comets 369
§ 102. General data on meteors 370
§ 103. Elements of the physical theory of meteors 373
§ 104. The results of photographic observations of meteors 375
§ 105. Spectra of meteors 379
§ 106. Radar observations of meteors 381
§ 107. The latest research techniques 386
§ 108. General data on the zodiacal light 387
§ 109. The nature of the zodiacal light 389

INDEX 393

PREFACE

This third volume of "A Course in Astrophysics and Stellar Astronomy" deals with observational results and their interpretation. It is not the aim of this book either to present all existing astrophysical theories and hypotheses or to discuss systematically general problems in the modern theories of cosmogony and cosmology. Such subjects have been treated in special textbooks and monographs.* However, in all the astronomical literature so far there has been no book which gives a systematic presentation of observational results together with methods for their interpretation. This fact has determined the nature and the scope of this text. It mainly presents facts, and the principal theoretical studies related to an interpretation of the observational data are given only briefly (they are often just mentioned in the references). As far as possible, we have tried to avoid controversial theoretical questions for which no definite answers have yet been found, since this book does not represent a survey of astrophysics and stellar astronomy in the sense of the volumes edited by Kuiper on the sun and the solar system. Rather, it constitutes a textbook for young astronomers, both graduates and undergraduates. For this reason, it was impossible to include a systematic discussion of general and complex problems related to cosmology and cosmogony, the proper place for which is in more specialized monographs.

For the teaching of astrophysics in a university, this book should be used in parallel with a course in theoretical astrophysics. Consequently, in our presentation it is assumed that the reader is already familiar with the rudiments of theory.

This volume, Volume III, is divided into two parts: 1) The Sun, and 2) The Planetary System (planets, comets, and meteors).

Part I mainly stresses so-called solar activity, that is, the processes occurring on the surface of the sun. These processes are inevitably interrelated, since they represent different stages in the development of large-scale motions on the solar surface. These motions produce so-called active regions, which in a certain sense may be compared with terrestrial cyclones and anticyclones.

The study of solar activity has a direct bearing on various geophysical problems: on changes in the magnetic field and ionosphere of the earth, on the influence which solar phenomena have on weather and climate, and on other related questions. All these problems, which until fairly recently remained largely subjects of statistical speculation, are now being put upon a sound physical foundation; this has been possible because, during the last two decades, solar physics, thanks to the introduction of new research

* Ambartsumyan, V.A., E.R. Mustel', A.B. Severnyi, and V.V. Sobolev. Teoreticheskaya astrofizika (Theoretical Astrophysics). — Gostekhizdat. 1952 [English translation published by Pergamon Press].

methods, has at last begun to justify its name. At present, spectrophotometric and spectroheliometric investigations and radio-astronomical data make possible a detailed study of the physical conditions on the solar surface. In addition, it is now possible to determine, for individual solar formations, the atomic concentrations of different elements in different stages of ionization and excitation, the concentrations of free electrons, the temperature, the magnetic-field strength, and various other physical quantities. We are therefore in a position to offer a quantitative description of the processes occurring on the surface of the sun and to explain the influence which these processes have on terrestrial phenomena, both of which subjects require a familiarity with the main properties of the ultraviolet-radiation fluxes and corpuscular streams ejected by active solar regions.

Unfortunately, the relation between solar activity and phenomena in the earth's atmosphere and in the geomagnetic field, as well as the influence of the cyclic solar activity on climate, and other earth-sun problems too, could not be included in this text, since to do so it would have been necessary to ask a group of geophysicists to participate in the preparation of this volume.

Some very basic changes have also occurred in the study of the planetary system. Especially, many new data are now available on comets, which by virtue of their structure may be regarded as unusual objects in the solar system. In chemical composition, comets are not unlike the major planets, whereas they in no way resemble the asteroids, which until recently were regarded as bodies related to comets. New data have also been obtained on the atmospheres of the terrestrial planets and on the properties of their surface structure. A radical change has taken place in our conceptions of the major planets, whose internal structures place them in an intermediate category between the stars and the terrestrial planets. Consequently, the part of the text dealing with the solar system has been revised considerably and expanded, in comparison with the 1936 edition.

Although the authors have endeavored to include in this work all the advances which have been made to date in astrophysics, still this field develops so rapidly that it is impossible to tell whether, after 10 or 15 years, a substantial portion of the material in this book will not have to be revised in turn and a new edition prepared.

The Editorial Board

PART I
THE SUN

V. P. VYAZANITSYN

Chapter I

INTRODUCTION

N 67-15485

The sun is an ordinary member of the stellar universe, a typical star at a certain stage of stellar evolution. At the same time, it is the central body of the planetary system to which the earth belongs. The physical study of the sun (solar physics) is one of the most important and comprehensive fields of astrophysics. In a certain sense, an "accidental" factor, namely the nearness of the earth to the sun, is in the final analysis the only reason for the special significance attached to the study of this star.

The high intensity of the solar radiation received by an observer enables him to use instruments with maximum resolving power and dispersion for spectroscopic investigations of the sun, instruments which are inapplicable in stellar spectroscopy. Moreover, the sun is the only star which has observable angular dimensions. All this is conducive to an especially detailed and thoroughgoing study of the structure and physical state of the upper solar layers. However, the results obtained for the sun may also, with certain reservations, be applied to other stars as well and thus serve as a basis for the development of astrophysics as a whole.

Astrophysical research is quite valuable from the viewpoint of general physics, since it is concerned with the properties of matter which often exists under conditions which cannot yet be reproduced in terrestrial laboratories. In this respect, solar physics contributes abundant data of great variety, and in the past it has been of considerable assistance in the development of various branches of physics. It suffices to mention, as an example, the role played by solar research in the development of spectroscopic analysis. However, our understanding of the physical conditions under which matter exists on the sun is in many cases still not thorough enough, as is evident from the difficulties involved in the interpretation of certain observed phenomena within the framework of modern solar physics.

Finally, solar research is of great importance in geophysics, since the conditions at the earth's surface and in the terrestrial atmosphere are determined by the flux of energy received from the sun. Solar activity (sunspots, flocculi, chromospheric flares) causes such terrestrial phenomena as magnetic storms and polar auroras, and it also affects radio-communication conditions. The study of the dependence of terrestrial phenomena on solar phenomena is one of the most important aspects of contemporary solar physics.

§1. General data. The main distance measure within the solar system is the astronomical unit, which is equal to the average distance a between the earth and the sun. If the solar parallax is taken as $p = 8''.7984$ (Rabe, 1950), this apparently being the most exact figure available, and if

the equatorial radius of the earth is taken as R_δ = 6378.2 km (Krasovskii), then a is given by the relation

$$a = \frac{206{,}265''}{p} R_\delta = 1.4953 \cdot 10^{13} \text{ cm} \approx 150 \cdot 10^6$$

Because of the ellipticity of the earth's orbit, this distance varies annually by 1.6% in either direction, that is, by 2.5 million km. Light covers a distance equal to a in 498 sec. As a comparison, let us recall that the annual parallax of the star nearest to the solar system (α Centauri) is only 0".75, which corresponds to a distance of 4.3 light years, or about 270,000 astronomical units.

The angular diameter of the sun varies from 32'31" at perihelion to 31'27" at aphelion, and has a value of 31'59".26 at the average distance from the earth. The last value, obtained by Auwers (1891) from many series of heliometric measurements, is the one accepted in all astronomical almanacs. Given the distance and the corresponding angular size of the sun, we find its linear radius to be

$$R_\odot = 695{,}700 \text{ km},$$

which is 109.1 times the radius of the earth. A comparison of the linear and angular dimensions of the sun shows that 1" on the solar surface corresponds to about 725 km.

The solar diameter is 1/107.4 of the average distance between the earth and the sun. This is, incidentally, the basis for the useful practical rule that the diameter of the solar image at a telescope focus, expressed in centimeters, is approximately equal to the focal length of the telescope, expressed in meters.

Many authors, starting with Secchi and Rosa in the last century and including such contemporary names as Cimino (1944), Meyermann (1950), and Gianucci (1955), have concluded, on the basis of a long series of measurements, that the solar diameter is subject to slight periodic variations which are in phase with the 11-year (or 22-year) cycle of solar activity. Different authors have obtained different amplitudes for the diameter fluctuation, ranging from several arc seconds down to one or two tenths of an arc second, and sometimes one or more secondary periods have also been detected. However, the actual existence of this phenomenon is still open to question, due to the considerable systematic errors in the relevant measurements. At any rate, the fluctuation amplitude for the solar diameter, if there is any variation at all, should be substantially less than 1'. The difference between the equatorial and polar diameters of the sun reported by some authors is also uncertain. Although the accuracy in these measurements is higher, because the difference in the diameters rather than the diameters themselves is measured, still the probable measurement error is comparable with, or even greater than, the difference detected (some tenths or hundredths of an arc second). Waldmeier found that, if at all detectable, the difference in diameters characterizing the solar oblateness cannot exceed 0".01, which corresponds to $\sim$7 km.

The solar mass, according to the data of celestial mechanics, is 332.488 earth masses (Rabe, 1950). Taking the mass of the earth as $5.975 \cdot 10^{27}$ g (Birge, 1942), we find that the solar mass is $1.9866 \cdot 10^{33}$ g. Thus, taking

into account the dimensions of the two bodies, we find that the average density of the sun is 0.256 of that of the earth, namely 1.41 $g \cdot cm^{-3}$. On the basis of the same initial data, the acceleration of gravity at the surface of the sun is found to be 273.8 m/sec^2, which is 28 times that at the earth's surface.

Various authors have compared the brightness of the sun with that of the stars. After an analysis of various measurements, Russell (1916) established that the visual magnitude of the sun is $-26^m.72$ and that the photographic magnitude is $-25^m.93$. Somewhat more recent researches (Kuiper, 1938; Woolley and Gascoigne, 1948; de Vaucouleurs, 1949) have put the average photovisual magnitude of the sun at $-26^m.88$. From Russell's figures it follows that the absolute visual magnitude of the sun is $+4^m.85$, whereas its color index is $+0^m.79$; thus, the sun is a yellow star located in the main sequence of the spectrum-luminosity diagram. More precisely, spectrally the sun belongs to the class of dG2 stars, which have temperatures of $\sim 6000°$.

§2. The total solar radiation. The study of the physical nature of the sun and stars is based on an analysis of the solar and stellar radiation reaching the earth. All the properties of the radiation emitted by a source (blackbody) in a state of thermodynamic equilibrium are expressed by the familiar laws of radiation (the Stefan-Boltzmann law, Planck's law, and others) and are determined by a single parameter, the temperature. However, the radiation of the sun deviates from that of a perfect blackbody, as is evident, for example, from the simple fact that the solar disk is not uniformly bright, becoming darker from the center to the limb (radiation nonisotropy). Different temperatures will thus correspond to different properties of the solar radiation, and the only correct procedure in this case is to obtain the various characteristics of the radiation directly from observations.

In order to form a complete picture of solar radiation, it suffices to study the absolute distribution of energy in the spectrum of the solar disk (either the entire disk or some point on it) and the decrease in disk brightness from center to limb at various wavelengths. In practice, the solution of the first problem is connected with the determination of the solar constant (see Volume I, Chapter XXXI), which characterizes the total radiation of the sun and is a very important parameter in meteorology.

The procedure currently used to determine the solar constant has been described elsewhere (Vol. I, Chapter XXXI). It consists in combining pyrheliometric measurements of the total solar energy at various zenith distances with spectrobolometric measurements of the relative distribution of energy in the solar spectrum. An extrapolation beyond the terrestrial atmosphere of the spectrobolograms for separate points, using Bouguer's formula, gives the true relative distribution of energy in the spectrum, while a comparison of the spectrobologram areas with the corresponding pyrheliometer readings makes it possible to express the intensity in absolute units. The area of the final curve obtained after the appropriate reductions have been made (reduction to average earth-sun distance, corrections for infrared absorption bands of water vapor and for spectrum edges cut off by the atmosphere) gives the solar constant.

Measurements of the solar constant have been made by many researchers. Those which are of greatest significance, both with regard to instrument

accuracy and measuring technique and with regard to the observational data obtained, are the studies carried out by the Smithsonian Institution (Abbot and co-workers) beginning in 1903 and lasting through several decades. The standard value of the solar constant obtained from Abbot's observations is

$$1.94 \text{ cal/cm}^2 \text{ min.}$$

However, the absolute scale on which Abbot's measurements are expressed ("the revised Smithsonian scale" of 1913) is apparently too high, as his latest observations indicate. After the introduction of an appropriate correction, first computed by Unsöld (1938), the solar constant is found to be

$$1.90 \text{ cal/cm}^2 \text{ min.}$$

Nevertheless, the solar constant can be specified at present only within accuracy limits of several percent, mainly due to the insufficient precision of the absolute intensity measurements in the infrared and ultraviolet regions of the spectrum. Discussions of these corrections in recent publications have indicated that Abbot's solar constant may be somewhat low. For example, according to Allen (1950) the solar constant should be increased to 1.970, while Nicolet (1951) gives a value of 1.98. There are reports that rocket measurements of the solar constant, carried out using bolometers equipped with tracking systems which continually point them toward the sun, give a value of about 2.00 cal/cm^2 min at an altitude of around 65 km. A determination of the solar constant using the ordinary method of spectral-curve integration, but using rocket data on the short-wave region of the solar spectrum, also gave a value of $2.00 + 0.04$ cal/cm^2 min.

It is easy to calculate that a solar constant of 1.90 cal/cm^2 min corresponds to a total solar-radiation flux (the radiation emitted from 1 cm^2 of solar surface per second in all directions) of $6.121 \cdot 10^{10}$ erg/cm^2 sec. Consequently, applying the Stefan-Boltzmann law,* we find that the effective temperature of the sun is

$$T = 5732^\circ\text{K}.$$

A solar constant of 2.00 cal/cm^2 min corresponds to an effective temperature of 5806°K.

§3. Fluctuations in solar radiation. The question of whether or not the amount of energy emitted by the sun varies is of considerable importance both in solar physics and in meteorology.

Two methods have been used to study possible fluctuations in solar radiation: the actinometric method, according to which direct measurements of the solar constant are made, and the astrophysical method, based on comparing the apparent brightness of the planets illuminated by the sun with the brightness of stars. The astrophysical method is superior to the actinometric method, in that it is independent of variations in the transparency of the terrestrial atmosphere. However, on the other hand, the astrophysical method does not allow for temporary turbidity and other changes in the planetary atmospheres, which may introduce analogous systematic errors into the results. The series of photoelectric measurements of planetary brightness obtained by Guthnick and Prager in 1917 and

* The most recent value of the Stefan-Boltzmann constant is $\sigma = 5.6716 \cdot 10^{-5}$ erg cm^{-2} deg^{-4} sec^{-1}.

1921 at the Babelsberg Observatory (for Jupiter and Saturn), the measurements of Stebbins (for the satellites of Jupiter), and also those of Stebbins and Jacobson in 1927 at the Lick Observatory (for Uranus), have shown very definitely that, if the solar brightness fluctuates at all, then the variation is less than 1%. The reason for the long-period fluctuations in planetary brightness is still not quite clear, since no definite correlation has been established as yet between these fluctuations and solar activity or between the fluctuations for different planets.

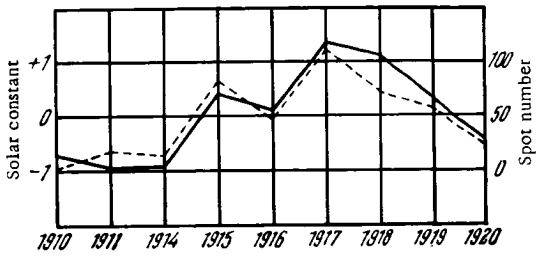


FIGURE 1

Although the accuracies of relative measurements of the solar constant are at least ten times higher than those of absolute measurements, the direct effect which the variable atmospheric transparency has on the measurement results greatly complicates the problem of detecting actual fluctuations of the solar constant. This effect can be eliminated by a statistical comparison of the solar-constant values obtained simultaneously at different stations widely removed from one another. Such comparisons have been carried out by several authors, and the results have invariably indicated a positive correlation between the different series of measurements. For example, Kron obtained a correlation coefficient of +0.51 for measurements made at Mount Wilson and at Algiers in 1911 – 1912. Bernheimer (Smithsonian-Institution stations, 1920 and 1921 – 1922) obtained coefficients of +0.29 and +0.34, Kimball (same stations, 1922 – 1924) obtained +0.18, etc. These results apparently indicate that fluctuations in the solar constant do exist. However, it should also be kept in mind that sometimes a temporary turbidity of the earth's atmosphere may be present over enormous territories. For example, after the eruption of Mt. Katmai in Alaska in 1912, a reduced atmospheric transparency due to the presence of volcanic dust was observed for over a year at stations in America, Europe, and Africa. Consequently, even comparatively high correlation coefficients cannot be regarded as adequate proof of the existence of solar-constant fluctuations, since they may to a large degree be due to variations in the state of the earth's atmosphere. This is directly confirmed by the studies of Bernheimer and Marvin, who detected a correlation between the value of the solar constant and the coefficient of atmospheric transparency.

In order to check the existence of the long-period fluctuations of the solar constant, Abbot in 1926 proposed the method of "selected pyrheliometry," according to which measurements made in different years are compared under conditions which are as close as possible (during the same month, on days of equal transparency and humidity, at equal zenith distances, etc.). The resulting curve [solid line in Figure 1] shows a very close fit with the solar-activity curve (dashed line in Figure 1).

Ångström obtained a coefficient of +0.63 for the correlation between the yearly average values of the solar constant (according to the above-mentioned observations at Mount Wilson, 1910 — 1920) and the solar activity (Wolf numbers); thus, the agreement between the two was fairly good. However, the results of analogous comparisons carried out by various authors have turned out to be very contradictory. For example, according to Waldmeier the solar constant (solid line in Figure 2) for the years 1920—1939 showed no definite correlation with the solar activity (dashed line in Figure 2). On the other hand, Aldrich and Hoover, on the basis of the data of the Smithsonian-Institution stations for 1923 to 1952, concluded that the solar constant varies in accordance with the sunspot number.

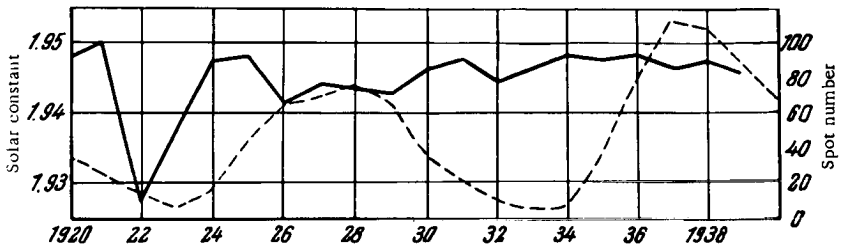


FIGURE 2

On the whole, it can only be stated that the question of solar-constant fluctuations is still insufficiently clear and that further research is required. The fluctuations observed are undoubtedly determined to a certain extent by the variations in the earth's atmosphere, and it may be that this factor is entirely responsible for the short-period fluctuations. The existence of the long-period fluctuations appears to be quite plausible, although as yet no conclusive proof has been produced.

In apparent contradiction to the minor variations in the solar constant, large-amplitude fluctuations in the intensity of the ultraviolet radiation of the sun are observed. However, as early as 1923 — 1925 it was established that the intensity of short-wave radiation is subject to considerably larger changes than that of long-wave radiation; also, more recent studies at the Smithsonion Institution (Abbot and others) have shown that on the whole the variations decrease sharply with wavelength.

Systematic measurements of the ultraviolet radiation were carried out by Pettit at Mount Wilson from 1925 to 1938. He measured the ratio of the intensity of the ultraviolet radiation in a small spectral region near λ = 3200 Å, which is outside the ozone-absorption band, to the intensity of the green emission near λ = 5000 Å (this ratio was arbitrarily set at unity for June 1924). Pettit's observations showed that the ultraviolet radiation of the sun fluctuates by as much as 20% or more, that is, its fluctuations are about ten times those of the solar constant. However, Bernheimer has shown that these fluctuations are to a large degree due to the terrestrial atmosphere. In addition to the annual variation in ultraviolet-radiation intensity, attributed to the variation in atmospheric transparency, there is also a distinct long-period fluctuation which brings about a change from year to year. A comparison of the intensity curve of the ultraviolet radiation for 1925 through 1938 with the Wolf-number curve for the sunspots (Figure 3) shows a close correspondence up to 1931, after which the

radiation curve maintains a fairly constant level in spite of the increase in solar activity. Waldmeier cites the reduction in the sensitivity of the measuring apparatus by 0.55 units from 1925 to 1938 as a possible reason for the observed noncorrespondence of the two curves.

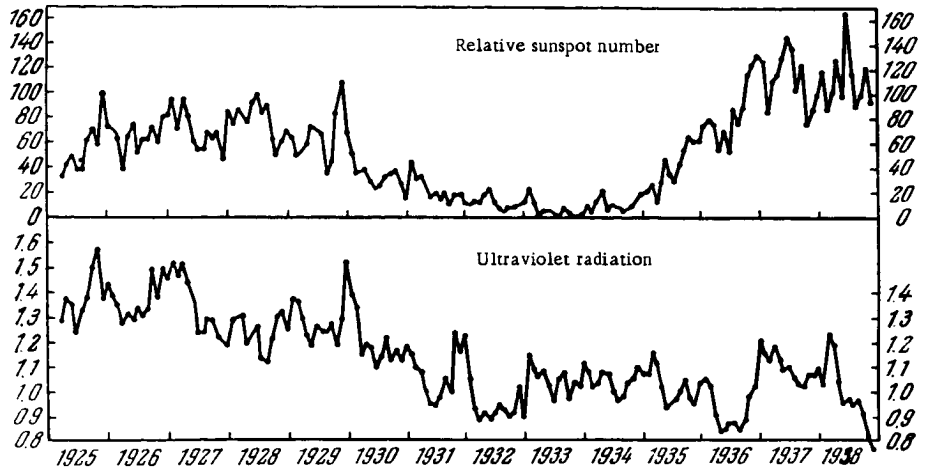


FIGURE 3

§4. The energy distribution in the solar spectrum. The main difficulty encountered in determining the energy distribution in the continuous solar spectrum is how to take into account the distortions introduced by partial absorption of the solar energy in the atmospheres of the earth and the sun. The continuous absorption in the terrestrial atmosphere (extinction) produces an overall weakening of the solar spectrum, with an increase in the degree of attenuation from red to violet. In addition, there are present numerous lines and molecular bands, both telluric (oxygen, water vapor, carbon dioxide in the red and infrared regions, and ozone in the ultraviolet region) and solar in origin. The lines become more frequent in the violet region, so that in the ultraviolet it is actually difficult to detect the line-free continuous background.

In order to allow for extinction in the earth's atmosphere, the intensity of the solar spectrum is always measured, as mentioned in §2, for various zenith distances of the sun, after which the measured intensities are extrapolated beyond the atmosphere, using Bouguer's formula. When comparing the measurement results of different observers, we must distinguish clearly whether the results refer to: 1) either the undistorted continuous spectrum between the lines or the diffuse-line spectrum obtained with low-resolution instruments and 2) either the spectrum of the integral solar radiation (that is, the spectrum of the radiation emitted by the entire solar disk) or the spectrum of some particular region (usually the center) of the disk. A reduction for attenuation of the spectrum by absorption lines or for the limb darkening of the solar disk is introduced in the respective cases.

After Lamansky s initial measurements in 1871, many researchers have continued the study of the energy distribution in the solar spectrum. Table 1 gives the intensities of the integral solar radiation (in erg $cm^2 sec^{-1}$ over

1 Å of the spectrum and per unit solid angle) according to Abbot's measurements for 1920 — 1922, reduced to absolute units by Minnaert (for a solar constant equal to 1.946 cal/cm^2 min); in the region of λ from 15,000 to 22,000 Å, Peyturaux's corrections have been introduced. These intensities, which are not corrected for the attenuation of the spectrum by Fraunhofer lines, characterize the solar radiation reaching the planets.

TABLE 1

3415 Å	$14.4 \cdot 10^5$	5980 Å	$26.8 \cdot 10^5$	13020 Å	$6.0 \cdot 10^5$
3504	15.5	6238	25.2	13770	5.2
3600	16.4	6530	23.8	14520	4.6
3709	17.5	6858	22.6	15280	4.0
3853	16.6	7222	20.2	16030	3.5
3974	18.8	7644	17.8	16700	3.1
4127	26.5	8120	15.6	17380	2.6
4307	26.4	8634	13.5	18700	2.0
4516	30.2	9220	11.6	20000	1.5
4753	31.2	9861	10.4	21230	1.2
5026	30.1	10620	9.3	22420	0.9
5348	27.9	11460	7.9	23480	1.1
5742	27.0	12250	7.0		

In order to correct the energy-distribution curve of the solar spectrum for the Fraunhofer lines, we should determine the total absorption in the lines for small successive spectral regions. This may be done, for example, using the calibration of Rowland estimates of line intensities (see Chapter II) or the Utrecht photometric atlas of the solar spectrum. Such a study was carried out in 1934 by Mulders, who found that on the whole the lines absorb 8.3% of the solar radiation, although the correction is far from uniform for different spectral regions (it increases from $\sim$1% near λ = 8000 Å to $\sim$24% near λ = 3500 Å). Table 2 gives the energy distribution in the spectrum for the center of the solar disk obtained by Minnaert, who combined Abbot's measurements corrected by Mulders (and of course reduced to the disk center) with Peyturaux's measurements.

TABLE 2

3200 Å	$24.3 \cdot 10^5$	5600 Å	$36.3 \cdot 10^5$	11000 Å	$10.9 \cdot 10^5$
3400	28.2	5800	34.7	12000	7.8
3600$^-$	31.8	6000	33.1	13000	6.6
3700$^-$	33.5	6200	31.6	14000	5.4
3700$^+$	45.2	6400	30.0	15000	4.5
3800	45.6	6600	28.5	16000	3.7
4000	46.0	6800	27.1	17000	3.0
4200	45.9	7000	25.6	18000	2.5
4400	45.3	7500	22.4	19000	2.0
4600	44.3	8000	19.5	20000	1.7
4800	43.2	8500	17.2	21000	1.4
5000	41.6	9000	15.2	22000	1.2
5200	39.8	9500	13.6	23000	1.0
5400	38.0	10000	12.1		

The results of the two authors show considerable discrepancy, especially in the ultraviolet and infrared regions, and further research is clearly necessary.

Figure 4 gives the spectral curves of the energy distribution for the entire solar disk (a) and for the center of the disk (b), according to Tables

1 and 2. It is interesting to note how rapidly the intensity of the spectrum drops in the short-wave region; another remarkable feature is the "Balmer discontinuity" (λ = 3647 Å). For comparison purposes, Planck's curve (c) for 6000 °K is also given.

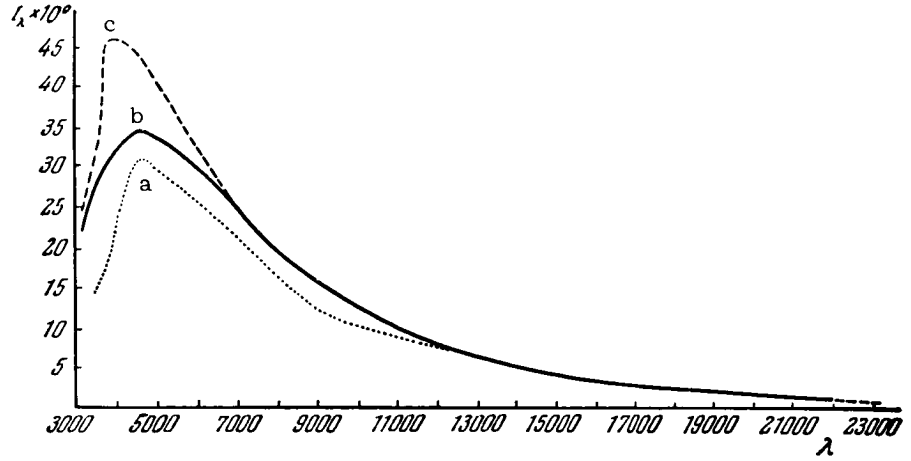


FIGURE 4

The energy distribution in the continuous spectrum of the sun cannot be described exactly by using Planck's formula, and consequently it cannot be characterized by a single color temperature for the entire region observed. Thus, different color temperatures are computed for individual limited regions of the spectrum. For example, Brill found the following color temperatures for the integral solar radiation: 5835°(3100 – 5070 Å), 6075° (4200 – 7000 Å), and 5775° (infrared region), while Woolley gives 6500° (3800 – 6700 Å).

The spectrum changes somewhat from the center of the solar disk toward the limb. According to Minnaert, the integral solar spectrum is identical with the spectrum of the disk at a distance 0.749 to 0.775 of a solar radius from the center of the disk. The temperatures obtained for the central parts of the disk are therefore considerably higher. Brill found Planck temperatures at the center of the disk of 6190°, 6435°, and 6075° for the photographic, visual, and infrared regions, respectively; Canavaggia and Chalonge obtained values of 6500° (3300 – 3600 Å) and 7500° (3700 – 5000 Å).

TABLE 3

Distance	λ_{max}, Å	Distance	λ_{max}, Å
0.00	4580	0.75	4780
0.20	4630	0.825	4830
0.40	4670	0.875	4890
0.55	4710	0.92	4960
0.65	4740	0.95	5050

The temperature of the sun can also be calculated spectrally using Wien's law: $\lambda_{max} T$ = 0.2880 cm deg. The values of λ_{max} for spectra at various

distances from the center of the disk, according to Abbot's measurements, are given in Table 3. However, since the position of the energy maximum in the solar spectrum has been determined only very inaccurately, these values of T cannot be taken so seriously.

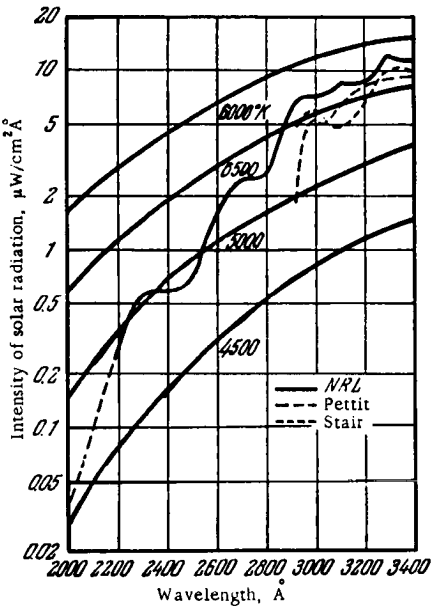


FIGURE 5

One of the most important tasks of the solar physicist is to widen the portion of the solar spectrum which can be studied, and some important achievements have already been made in this direction. The ozone bands (the equivalent ozone layer for the atmosphere is 2 to 3 mm thick) occur in the solar spectrum in the ultraviolet, near λ = 2900 Å. The attempts of various researchers to penetrate beyond this limit by making observations on high mountains and by using sounding balloons did not on the whole meet with spectacular success. It was only in the postwar years, due to the development of rocket technology, and due to the high-altitude ascents of rockets carrying measuring apparatus, that a rapid advancement was made into the short-wave region of the solar spectrum. Starting in 1946, American scientists launched many V-2 (170 km), Aerobee (110 km), and Viking (220 km) rockets. Spectrographs with high light-gathering power and other instruments with wide fields of view were designed, and special measures were taken to stabilize the Viking rockets. Also, special stabilizing devices were developed enabling the instruments to track the sun by means of servosystems. The spectral region from 3000 to 1900 Å was given a first-order study using concave-grating spectrographs (600 lines per mm and radius of curvature from 40 to 60 cm) with slits and without slits. Due to the difficulties involved in achieving long exposures with rocket-borne instruments, the extreme ultraviolet, beyond λ = 1900 Å, was studied using nonspectroscopic methods (thermoluminescent phosphors

and photon counters connected to various filters were used), and the results were telemetered back. In 1948, X-rays were recorded photographically through the beryllium windows of a V-2 rocket; the X-rays were then photographed repeatedly onto beryllium-coated (< 7 or 8 Å) or aluminum-coated (8 to 15 Å) film. Thermoluminescent phosphors and photon counters were also used in conjunction with beryllium or aluminum windows. These studies are still in progress.

Figure 5 shows the intensity of distribution in the solar spectrum from λ = 3400 Å to λ = 2000 Å, as obtained in 1953 at the Naval Research Laboratory (USA) from measurements of photographs taken using rockets. The relative intensities down to λ = 2200 Å were based on measurements of the angular-arc spectrum, and then the curve was extrapolated down to λ = 2000 Å. The absolute scale was derived from measurements of the energy distribution throughout the entire solar spectrum, taking into account a new revision of the solar constant. The curve shows that the intensity of the solar spectrum decreases between λ = 3400 Å and λ = 2000 Å, relative to the Planck curve for T = 6000°. Near λ = 2200 Å the solar radiation corresponds to a temperature of 4900°.

The application of photography to the study of the infrared region of the solar spectrum is limited to the range below λ = 13,000 Å = 1.3μ. PbS cells, which have been used predominantly for infrared research in recent years, raise this limit to 3.5μ, but the instrument which actually reaches farthest into the infrared is the bolometer. Nevertheless, the existence of solar radiation has so far been reliably verified only up to $\sim$24μ.

The measurable solar spectrum was widened considerably with the discovery of solar radio emission. This high-frequency radiation, which reaches the terrestrial observer in the wavelength interval between several millimeters and 12 m (the lower shielding limit of the earth's atmosphere), originates in the upper layer of the solar atmosphere (the chromosphere and corona).

§5. Limb darkening. Any photograph of the sun will show that the solar disk is not uniformly bright, and that the brightness decreases noticeably toward the limb (limb darkening).

In Figure 6, let s be an element of the solar surface at the center of the disk and let s' be an analogous element at a distance r (or $\sin\vartheta = r/R$) from the center; sO is the direction of the observer. Since the sun is spherically symmetric, the intensity of the radiation emitted by element s' in the direction O' is obviously equal to the intensity of the radiation from element s in the direction F, which is at an angle ϑ to the normal. Thus it follows that the radial variation of the solar-disk brightness characterizes the angular distribution of the radiation intensity of a solar-surface element.

Limb darkening arises because, as we look at points further from the center of the solar disk, the line of sight cuts through the upper layers of the sun more and more obliquely. Consequently, the relative amount of radiation from the deeper, hot layers decreases due to the increased absorption in the outer, cooler layers. A quantitative study of limb darkening is essential for the derivation of the temperature gradient of the photosphere and for the construction of photospheric models.

The radial intensity distribution of the solar disk in integrated (polychromatic) light was already studied by various scientists during the last century. A comparison of different series of measurements indicates that substantial random and systematic errors exist, and of

these the systematic error introduced by light scattering in the atmosphere and in the measuring instrument is quite important. Julius observed that limb-darkening measurements should be made during total or annular eclipses, so as to reduce this error to a minimum. His measurements during the eclipses of 1905 and 1912 actually gave a steeper drop in solar-disk brightness toward the limb. However, Julius's method is not free of systematic errors, either. The transit of the lunar shadow is accompanied by a drop in temperature, and consequently by a drop in atmospheric absorptivity as well, so that the emission curve for the eclipsed sun is nonsymmetric relative to the eclipse center [midtotality].

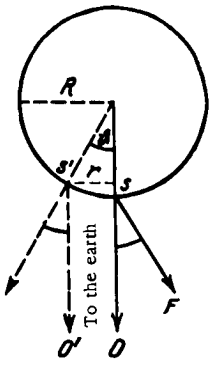


FIGURE 6

It should be noted too that when the limb darkening is measured directly using integrated light (for example, with a thermoelement), the results for the various spectral regions have incorrect "weighting factors," due to the variation of atmospheric absorption as a function of wavelength. Therefore, it is better to calculate the limb darkening in white light from measurements of the darkening at individual wavelengths, by taking into account the energy-distribution curve for the solar spectrum. The results of Milne (1922), obtained by this means using Abbot's measurements, are given in Table 4, where $I(\vartheta)$ denotes the radiation intensity at a distance $\sin\vartheta$ from the center of the disk.

TABLE 4

$\sin\vartheta$	0.05	0.40	0.55	0.65	0.75	0.825	0.875	0.92	0.95
$I(\vartheta)/I(0)$	1.000	0.955	0.912	0.871	0.822	0.769	0.722	0.685	0.612

Studies of the luminosity distribution over the solar disk at various wavelengths commenced about 70 years ago, and since then numerous measurements have been made over a wide region of the spectrum, using a diversity of methods. These studies showed that the limb darkening of the solar disk is a function of wavelength, and that the brightness attenuation increases for shorter wavelengths. The most complete results obtained were those of Abbot (and co-workers), who made systematic measurements over several years. The procedure used consisted in shifting the solar image across the slit of a spectrobolometer or a monochromator connected to a thermopile or photocell. Some observers have conducted measurements in monochromatic light using the above-mentioned method of Julius (the first to do this was Schwarzschild in 1903).

Figure 7 gives the curves for the limb darkening $I_\lambda(\vartheta)/I_\lambda(0)$ as a function of the wavelength λ, for various points on the solar disk defined in terms of the distance $\sin\vartheta$ (Minnaert). The corresponding data are given in Table 5. Table 5 and Figure 7 illustrate with satisfactory accuracy the nature of the darkening toward the solar limb over a large spectral region. It is interesting to note that there is a Balmer discontinuity near $\lambda = 3700$ Å, the magnitude of which agrees with the theoretical prediction.

TABLE 5

λ, Å	sin φ									
	0.000	0.200	0.400	0.550	0.650	0.750	0.825	0.875	0.920	0.950
3220	1	0.960	0.897	0.835	0.775	0.690	0.600	0.530	0.452	0.382
3700	1	0.984	0.934	0.871	0.811	0.730	0.652	0.580	0.499	0.432
4000	1	0.982	0.929	0.863	0.802	0.720	0.645	0.566	0.495	0.430
4500	1	0.984	0.939	0.883	0.827	0.755	0.682	0.610	0.538	0.474
5000	1	0.987	0.947	0.896	0.846	0.783	0.713	0.650	0.580	0.519
5500	1	0.989	0.953	0.907	0.861	0.803	0.740	0.684	0.618	0.560
6000	1	0.990	0.958	0.914	0.873	0.819	0.762	0.712	0.649	0.593
7000	1	0.992	0.966	0.929	0.894	0.847	0.797	0.754	0.698	0.647
8000	1	0.993	0.970	0.938	0.907	0.866	0.821	0.782	0.733	0.687
9000	1	0.994	0.974	0.946	0.920	0.884	0.846	0.811	0.766	0.724
10000	1	—	0.976	—	0.926	0.894	0.857	0.820	0.775	0.736
15000	1	—	0.984	—	0.952	0.927	0.902	0.875	0.840	0.807
20000	1	—	0.988	—	0.968	0.945	0.922	0.900	0.875	0.846
30000	1	—	0.990	—	0.973	0.958	0.944	0.930	0.911	0.889
60000	1	—	0.994	—	0.982	0.972	0.963	0.955	0.940	0.924
100000	1	—	0.997	—	0.992	0.986	0.979	0.974	0.963	0.953

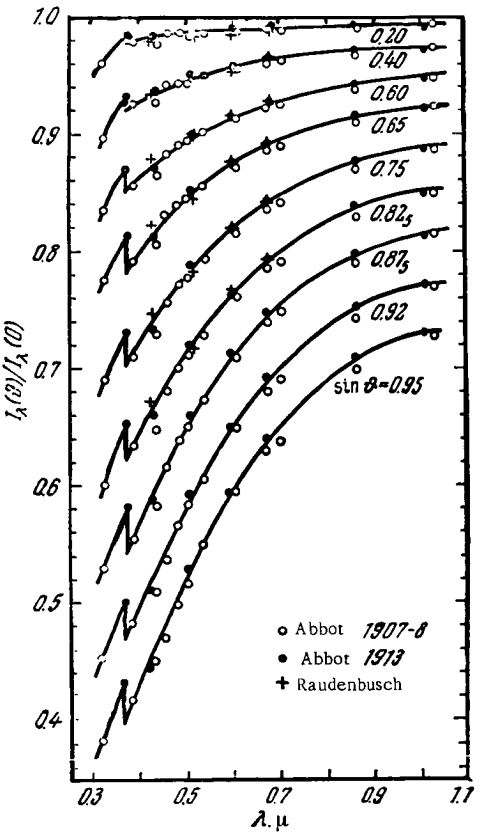


FIGURE 7

Also of considerable theoretical interest is the brightness variation of the disk just at the limb, but the corresponding measurements are exceedingly difficult due to limb fluctuations and to other effects. Such measurements can be made during total solar eclipses, prior to the onset of or immediately upon

completion of the total phase. Observations during eclipses have shown that the brightness decreases most rapidly just at the edge of the disk. For example, according to the measurements of Lindblad and Kristensen in 1945, the limb brightness decreases over 1 "by a factor of 100. This fact explains the apparent sharpness of the edge of the solar disk. The observations which have been made so far are still inadequate for establishing the exact law governing the brightness attenuation at the limb in various spectral regions.

TABLE 6

λ, Å	I_0/F	r	λ, Å	I_0/F	r	λ, Å	I_0/F	r
3230	1.487	0.758	5010	1.287	0.750	10310	1.130	—
3700	1.361	—	5340	1.265	0.751	12250	1.118	0.762
3860	1.407	0.749	6040	1.226	0.750	16550	1.088	0.774
4330	1.371	0.750	6700	1.201	0.758	20970	1.077	0.775
4580	1.326	0.752	6990	1.190	0.751			
4810	1.302	0.753	8660	1.158	0.763			

Table 6 gives the results obtained by Minnaert (1924) for the ratio of the intensity I_0 at the disk center to the average intensity F of the disk for various wavelengths; the distance r from the disk center at which the actual intensity is equal to the average intensity for the particular wavelength is also given. The table shows that the values of r for the entire spectrum lie within a very narrow interval, between 0.749 and 0.775 of the disk radius.

§6. Solar rotation. One of the most important solar phenomena is the rotation of the sun about its axis. The solar rotation is directly observable from the daily shifting of various formations on the solar disk (sunspots, faculae, flocculi, filaments). The measurement of the exact coordinates of these formations at successive times enables us to derive the rotation elements of the sun, that is, the constants which specify its rotation. Since all the solar formations also possess proper motions over the solar surface, it is possible to distinguish the rotation from the apparent motion only if we make some assumptions concerning the nature of the proper motions. For example, assuming that the proper motions are random, we can determine the solar rotation by measuring a sufficiently large number of individual formations whose proper motions can on the whole be taken as zero.

Because of the rotation, all the lines in the spectrum of the eastern solar limb are displaced toward the violet, relative to the positions of these lines for the disk center, whereas the spectral lines for the western limb are shifted toward the red. If the Doppler shift at a wavelength λ is $\Delta\lambda$ for some point on the sun at a distance ϑ from the central meridian, then it is easy to show that the linear velocity of rotation is

$$v = \frac{\Delta\lambda \cdot c}{\lambda \sin \vartheta \cos \varphi_0},$$

where φ_0 is the heliographic latitude of the disk center and c is the velocity of light. The measurement of the Doppler shift of the lines in the solar-limb spectrum provides us with an alternative means of investigating the solar rotation. The advantage of the spectroscopic method is its independence of transient and variable formations (sunspots, etc.) and also its

applicability to latitudes outside of the sunspot zone. The measurements should be made differentially, by simultaneously photographing the spectra of two diametrically opposite points on the solar limb side by side, since such a procedure doubles the apparent shift of the lines and moreover reduces the effects of various systematic errors.

The orientation of the solar rotation axis in space is specified by two elements, the inclination angle i between the solar-equator plane and the plane of the ecliptic and the longitude Ω of the ascending node of the equator. A third rotation element is the period τ of sidereal rotation about the axis (at the equator; see below) or the corresponding angular velocity $\omega = 360^\circ/\tau$ (generally in degrees per average solar day).

Various authors have determined the solar-rotation elements from measurements of the sunspot positions.* In 1928, by international agreement the Carrington rotation elements (see Vol. I, §201) were adopted as a standard; these elements, derived from visual sunspot observations, have the following values:

$$i = 7^\circ.25, \quad \Omega = 73^\circ.6667 + (T - 1850) \cdot 0^\circ.01396.$$
$$\tau = 25^d.380, \quad \omega = 14^\circ.1844.$$

The second term in the expression for Ω takes into account the precession for the date T (expressed in years). In present studies of the solar rotation, τ is the only one of the three rotation elements which is usually determined, as being of astrophysical significance, and i and Ω are invariably taken according to Carrington.

Obviously, the synodic-rotation period for the sun (the time interval between successive transits of a solar formation through the central meridian) $\sigma > \tau$, and the daily synodic rotation angle $360^\circ/\sigma$ is equal to ω minus the daily orbital motion of the earth, which is $360^\circ/365.26 = 0^\circ.9856$. Using Carrington's value for ω, we obtain $\sigma = 27^d.27$.

However, the sun does not rotate as a rigid body. Its angular rotation velocity is a function of the heliographic latitude, a fact which was noted by the earliest sunspot observers but first studied systematically by Carrington. The spectroscopic measurements of Dunér (1887 – 1889 and 1899 – 1901) not only verified the solar-rotation law established from sunspots for the ±45° zone but also indicated its validity outside this zone (up to ±75°). Much attention was paid to the study of solar rotation during the following decades. The rotation law was determined by many investigators, either from the various solar formations or else spectroscopically, using the different spectral lines.

Various formulas are used as empirical expressions for the velocity of solar rotation as a function of the heliographic latitude, the commonest of these being Faye's formula:

$$v = a - b \sin^2 \varphi,$$

where a is the equatorial velocity and b is a coefficient known as the equatorial acceleration. (An analogous formula describes the angular velocity.) At present this formula is used as a standard expression for various results related to solar-rotation studies. The average equatorial velocity of the

* For the derivation of the corresponding equations, see /1/.

rotating layer, which has been determined spectroscopically by many observers, is close to 2.00 km/sec (14°.38 per day), which is 1% less than the value obtained from sunspots. However, the measurements of different authors show a scattering of as much as 10%.

A comparison of the results of determining the solar-rotation law according to the various solar formations shows that the angular rotation velocity is not constant in all the layers, but increases with the height of the level (sunspots → flocculi → prominences). This result appears to be confirmed by spectroscopic measurements, which indicate an increase in rotation velocity for higher origin levels of the lines used in the measurements (metallic lines → Ca 4247,* H_α, Na D-line, and others → emission H_α and K_3). However, the conclusion that the rotation velocity increases with altitude is still not definite. Despite their high internal accuracy, the measurements contain large systematic errors, as is evident from the sizable discrepancies between the results of different authors. The question of the variation in rotation velocity with time (in particular, in phase with the 11-year solar-activity cycle) is also undecided as yet, for the same reason.

REFERENCES FOR CHAPTER I

1. WALDMEIER, M. Ergebnisse und Probleme der Sonnenforschung. — 2 Auflage, Leipzig. 1955.
2. KUIPER, G. (ed.). The Solar System. Vol.I, The Sun, Chap.III, Chap. I. §9. — Russian Translation from English, Edited by V.A.Krat. 1957.

* [Apparently the 4227 line is meant here.]

V. P. VYAZANITSYN

N 67-15486

Chapter II

THE SOLAR LINE SPECTRUM

§7. The normal solar spectrum. Spectrum atlases and line catalogs. The normal spectrum of the sun, that is, the spectrum of an undisturbed part of the solar disk or the spectrum of scattered sunlight, is an absorption spectrum. Against the background of a bright continuous spectrum a great number of thin, dark lines, called Fraunhofer lines, may be distinguished. These lines originate in the lowest layer of the solar atmosphere, due to the absorption of radiation with certain characteristic wavelengths by the atoms of various elements. The Fraunhofer lines in certain spectral regions (mainly in the ultraviolet) are so closely spaced that the small portions of undisturbed background are barely detectable between them. Wide bands, made up of groups of very fine lines, are also present, due to absorption of radiation by the molecules of chemical compounds. Many lines and molecular bands in the Fraunhofer spectrum are nonsolar in origin, and appear due to absorption in the earth's atmosphere (telluric lines and bands).

Studies of the continuous spectrum make it possible to determine the temperature of the sun, as was shown above (Chapter I). In addition, if the intensity distributions (over the solar disk) for frequencies of the continuous spectrum are measured, the theory of radiation equilibrium gives us information concerning the structure of the photospheric layers. An even more powerful tool for investigating the structure and physical state of the solar atmosphere is the Fraunhofer spectrum. The observed position, intensity, and profile of a Fraunhofer line reflect to a certain degree the effects of many factors characterizing the state of the atmosphere (temperature, pressure, diverse macroscopic motions, magnetic and electric fields, etc.), and they also depend upon the mechanism of the re-emission of different parts of the line in the atmosphere (absorption, coherent or incoherent scattering, etc.). A detailed study of these characteristics of the Fraunhofer lines thus yields a variety of data on the physical conditions in the solar atmosphere. One of the main problems which has been solved by Fraunhofer-spectrum study is the analysis of the chemical composition of the sun.

The absorption lines in the solar spectrum were discovered by Wollaston in 1802 and later rediscovered independently by Fraunhofer in 1814; Fraunhofer was the first to subject the solar spectrum to a detailed study. He recorded 674 lines, and his drawings represented the first actual atlas of the solar spectrum. Fraunhofer's letter notations for the strongest solar lines are in part still used.

However, the full significance of this discovery was grasped only in 1859, when Kirchhoff, on the basis of laboratory experiments, gave a qualitative

explanation of the origins of the various spectra (the continuous spectrum and the emission and absorption line spectra). He also compiled an atlas of the solar spectrum, in which some of the Fraunhofer lines were identified with lines from terrestrial light sources.

An atlas of the visual spectrum of the sun was also published by Ångström in 1863. Since he used a diffraction grating rather than a prism, Ångström obtained constant-dispersion spectra, so that the lines in his atlas could be arranged according to wavelength. He identified many of the Fraunhofer lines. Ångström's atlas was supplemented by the results of Cornu, who in 1874 and 1882 investigated the ultraviolet region of the solar spectrum.

In 1895–1897 Rowland published an excellent atlas, as well as the first catalog of the solar-spectrum lines. These remained unsurpassed for several decades and played an important role not only in the development of solar physics, but also of astrophysics in general. He carried out his own observations using a concave diffraction grating. Rowland's atlas, which is a photographically enlarged copy of the spectrograms in the $\lambda\lambda$2967–6953 region, includes about 20,000 lines and has a total length of 13 m. A comparison of the solar spectrograms with laboratory spectrograms of the chemical elements obtained using the same spectrograph enabled Rowland to identify about 1/3 of all the lines in the catalog.

At a convention of an international commission on solar research in 1905 it was decided to remeasure the wavelengths of the solar-spectrum lines, taking as a reference not the wavelength of the sodium D_1 line (λ 5896.156), as done by Rowland, but the red line of cadmium (λ 6438.3696), which had been measured interferometrically to an accuracy better than 0.001 Å. The choice of this line as the primary standard for the wavelength system is explained by the fact that, unlike most lines, even the strongest instruments cannot resolve it into components; the standard meter is now expressed in terms of the wavelength of this line as well (1,553,164.13 wavelengths, according to Michelson). As secondary standards (basic lines of 2nd order) were chosen the lines of iron (voltaic arc), which are separated from one another by 50 Å; the latter lines were related interferometrically to the cadmium line. In 1922 a series of neon lines was adopted as the secondary standard, and also tertiary and auxiliary standards were suggested, mainly the lines of neutral iron. The second-order lines were related by means of diffraction spectrographs to iron lines of 3rd order (numbering 383), at intervals of 5 to 10 Å and with a final error less than 0.006 Å. The totality of these standards comprises the international wavelength system. The measurement of a large number of solar lines with reference to these standards, a project which was carried out at the Allegheny Observatory in collaboration with the Bureau of Standards (Burns, Meggers, and Kiess) and at the Mount Wilson Observatory (St. John and Babcock), has made it possible to reduce the Rowland lines to the international system (the actual corrections range from −0.100 Å near λ = 3000 Å to −0.225 Å near λ = 6600 Å).

Nevertheless, for various reasons there is some discrepancy between the wavelengths of the solar lines reduced to the international system and the laboratory wavelengths for the vacuum-arc spectrum. For example, there is a Doppler shift of the lines in the spectrum of the solar limb relative to the lines for the center, due to the rotation of the sun about its axis. Incidentally, this effect is utilized in order to discriminate between the solar lines proper and the many telluric lines, particularly in the infrared. In

1917 Meggers found that even after solar rotation is allowed for there remains a certain residual "center-to-limb" shift, which (according to Hunter's measurements of high-dispersion spectra) can be as high as +0.005 Å. The red shift of the solar lines predicted by the theory of relativity (Einstein shift) amounts to +0.008 Å near $\lambda = 3800$ Å and increases to +0.014 Å near $\lambda = 6600$ Å, and on the whole this has been confirmed by St. John in his comparison of the wavelengths of the solar lines and of lines in the spectrum of a vacuum source. Minor deviations from the predicted values were attributed by St. John to differences in the levels of origin for the corresponding lines, these residual shifts increasing algebraically with the level height. By measuring the lines in the limb spectrum in the $\lambda\lambda 3600-6750$ region, St. John established that the residual shift exceeds the Einstein shift on the average by 0.0015 Å, and this constitutes the "limb effect." Using the iron lines, Evershed found residual limb effects of +0.0060 Å near $\lambda = 3932$ Å, +0.0050 Å near $\lambda = 4433$ Å, and +0.0119 Å near $\lambda = 6200$ Å; this led to the conclusion that the limb effect varies with wavelength, as also noted by Adams. The minor shifts (generally toward the red) in the solar lines after the Einstein shift has been taken into consideration, and also the limb effect, have not yet been satisfactorily explained. Thus, a further study of this problem on the basis of measurements of high-dispersion spectra, as well as a thorough consideration of all the complicating effects (turbulence in the solar atmosphere, the asymmetry or spreading of lines, etc.), are required.

A ten-year project dedicated to revision of Rowland's tables was undertaken at the Mount Wilson Observatory (St. John, Moore, Weaver, Adams, and Babcock), and it culminated in 1928 in the publication of the "Revision of Rowland's Tables of Solar Spectrum Wavelengths."

This catalog (The Revised Rowland) includes 21,835 lines in the region from $\lambda 2975.483$ to $\lambda 10,218.6$, for which the wavelengths and intensity estimates on the Rowland scale are given. The wavelengths are reduced to the international system, which is based on the standards adopted in 1922. In 1928 the International Astronomical Union adopted a number of improved standards (iron lines), and minor corrections should correspondingly be introduced in the wavelengths of the Revised Rowland (ranging from −0.0006 Å to −0.008 Å for the various spectral regions) in order to make them agree with the new standards. In the catalog 12,502 lines are identified (57% of all the lines listed). The catalog also gives the intensities of the lines in sunspot spectra relative to the line intensities in the normal solar spectrum. Finally, a temperature classification of the lines, a line classification according to profiles as a function of pressure, and the excitation potentials are also given.

In 1947 Babcock and Moore published a catalog including about 7400 lines in the $\lambda\lambda 6600-13,495$ range. These authors also extended the Rowland table into the ultraviolet; wavelength measurements and intensity estimates were made for 665 lines in the $\lambda\lambda 2935-3060$ range.

About 26,000 solar lines have been recorded so far in the $\lambda\lambda$ 2935 − 13,495 range. In addition to the wavelengths, visual estimates of the line intensities on the Rowland scale are also given, and these extend from very weak lines (0000 on the Rowland scale or −3 in the Revised Rowland) to the familiar strong lines which have letter notations (H_α, H_β, ... for hydrogen, H and K for calcium, D_1 and D_2 for sodium, etc.).

Although the method of estimating the intensities of Fraunhofer lines on the Rowland scale is very simple and convenient and is widely accepted, still it does not meet the requirements of modern spectroscopic solar research, and such intensity estimates must be replaced by more exact line characteristics. On the basis of spectrophotometric measurements, Allen determined the central intensities and equivalent widths of the Fraunhofer lines in the λλ 3924 – 11,830 range. Of special significance in this respect is the Utrecht photometric atlas of the solar spectrum published in 1940. This atlas describes the solar spectrum directly in terms of intensities and covers a range of λλ 3332 – 8771. The scale of the atlas is 2 cm / Å. In 1951, at the McMath-Hulbert Observatory an atlas of the infrared solar spectrum was published for λλ 8465 – 25,242.

The spectroscopic data, both solar and laboratory, which are available to date are now being utilized in a project whose aim is to extend and to supplement the Revised Rowland to cover the λλ 2935 – 13,495 range of the solar spectrum. The line wavelengths in this catalog will be reduced to the 1928 international scale, and the intensity estimates on the arbitrary scale will be replaced by the equivalent widths measured according to the Utrecht photometric atlas. Both the identification of atomic and molecular lines and the values of the excitation potentials will be re-evaluated as well.

§8. **The identification of Fraunhofer lines.** The primary problem in solar physics which can be solved with the aid of the normal solar spectrum is that of the qualitative chemical composition of the sun; this problem obviously reduces to an identification of the solar lines with the lines of chemical elements. The identification of the lines in solar and stellar spectra is one of the most difficult tasks of astrophysics, since a spectrum is determined by several physical conditions which, even in the case of the sun, are insufficiently understood and very difficult to reproduce in the laboratory.

The simplest identification procedure is to compare directly a photograph of the solar spectrum with a photograph of the spectrum of some chemical element. In this procedure not only coincidence of the wavelengths of the solar and laboratory lines is taken into consideration but also the relative intensities of these lines. If in the solar spectrum the wavelengths of weak lines coincide, but the strong lines of the element in question are absent, then the agreement is accidental. The permissible discrepancy between wavelengths depends on the nature of the lines, on their intensity, width, spreading, asymmetry, etc. The comparison should be made with laboratory spectra obtained under various physical conditions. The blending of lines which may belong to the same or to different elements introduces a further complication.

It is very important that on the sun, as in the laboratory, it is possible to study the spectra corresponding to various physical conditions, such as limb and disk spectra, and sunspot and chromosphere spectra. In the spectra of sunspots, the solar disk, and the chromosphere, many lines differ appreciably in intensity, although their wavelengths (after appropriate reductions) agree well with one another. These differences are analogous to those observed respectively in the laboratory spectra for an electric furnace, an arc, and a spark. A sunspot is about 1000° cooler than the disk, so that for a transition from the disk spectrum to the sunspot spectrum the lines of atoms with low ionization potentials (Ca, Sc, Ti, V, etc.) are strengthened, while, conversely, arc lines with high excitation potentials

and spark lines are weakened. This fact was discovered at the beginning of the present century and served as a stimulus for the ambitious program undertaken by King at Mount Wilson Observatory, designed to provide a temperature classification of the laboratory lines observed in electric furnaces and arcs at various temperatures (classes I through V). Hundreds of lines of molecular origin which are present in sunspot spectra but absent in the disk spectrum differ from the atomic lines due to the magnetic fields of the sunspots (Zeeman effect), and this provides a further check for many identifications.

Most rare-earth lines, due to their relatively low ionization potential, are observed only in a singly ionized state both on the disk and in sunspots. Menzel and Thackeray demonstrated that in the spectra recorded at the onset and end of a total solar eclipse the spark lines for the rare earths, contrary to other lines, do not show an abrupt reversal upon transition from the emission spectrum of the chromosphere beyond the solar limb to the absorption spectrum of the disk; these lines are visible in emission far into the disk. This remarkable feature of the rare-earth lines is very useful for their identification.

The development of ionization theory and quantum theory, which explain the mechanism by which spectra are formed, increased considerably the possibilities for identifying solar lines, in comparison with the possibilities using the "coincidence" method. Given the measured wavelengths, the estimated intensities, and the temperature classification of the lines in a laboratory spectrum, we can compare these with the energy levels predicted by the theory of atomic structure. These levels correspond to the configuration of the surrounding electrons and are expressible in spectroscopic terms, so that we can identify the lines in the given spectrum of an element in a certain ionization state.*

Multiplets are particularly important for the identification of solar lines, since the relative intensities and positions of the multiplet components must be the same in both solar and laboratory spectra. For example, by applying the theoretical intensities for multiplets, we can derive the component-line intensities from the total intensity of the blended line. In order to verify the presence of a certain element on the sun, we first look for the multiplets containing the strongest lines with the lowest ground-level excitation potentials. If these lines are accessible to observation but are not present in the solar spectrum, then we may conclude that the element in question is absent from the solar atmosphere. For many elements these lines lie in the $\lambda < 2950$ Å range, for which only low-dispersion rocket spectra are available so far.

Lines can often be discovered in the solar spectrum if their exact positions, as predicted from the energy levels, are known. More than 1250 Fe I lines have been found on the sun by this method and later confirmed in part by laboratory measurements. The conditions on the sun (low density and thick photosphere) are more favorable for the creation of these weak lines than conditions in a laboratory arc. For quite abundant elements, lines of higher excitation potential should also be observed in the solar spectrum; in this case as well the strongest lines of the corresponding multiplets in the accessible region are first sought.

* The U.S. Bureau of Standards has published a summary of all known energy levels, derived from an analysis of optical spectra.

As noted previously, of the 21,836 lines in the Revised Rowland, 57% were identified; also, 57 chemical elements were identified. Almost all the unidentified lines are very weak. Only a few of them have Rowland intensity estimates of more than 3, and 98% of the unidentified lines in the catalog lie in the intensity interval from 1 to −3. The elements corresponding to the strongest Fraunhofer lines of solar origin in the visual spectrum are listed in Table 1.

TABLE 1

Line, Å	Rowland intensity	Element	Line, Å	Rowland intensity	Element
3934	1000	Ca+	5167	15	Mg
3969	700	Ca+	5173	20	Mg
4102	40	H	5184	30	Mg
4227	40	Ca	5889	30	Na
4341	20	H	5896	20	Na
4862	30	H	6563	40	H

At present, of the 26,000 lines registered in the photographic region of the spectrum about 70% have been identified. 68 different elements have been found to exist on the sun (Table 2), although the certainty of the identification differs for different elements. For example, only a single line was available for the identification of A (the forbidden λ 5536 line in the coronal spectrum, identified by Edlén as A X), Cd, Au, and Th; the presence of Tb, Er, Ta, and other elements is still not certain.

TABLE 2

1 H	11 Na	21 Sc	30 Zn	42 Mo	51 Sb	64 Gd	73 Ta
2 He	12 Mg	22 Ti	31 Ga	44 Ru	56 Ba	65 Tb	74 W
3 Li	13 Al	23 V	32 Ge	45 Rh	57 La	66 Dy	76 Os
4 Be	14 Si	24 Cr	33 As	46 Pd	58 Ce	68 Er	77 Ir
5 B	15 P	25 Mn	37 Rb	47 Ag	59 Pr	69 Tu	78 Pt
6 C	16 S	26 Fe	38 Sr	48 Cd	60 Nd	70 Yb	79 Au
7 N	18 A	27 Co	39 Y	49 In	62 Sm	71 Lu	82 Pb
8 O	19 K	28 Ni	40 Zr	50 Sn	63 Eu	72 Hf	90 Th
9 F	20 Ca	29 Cu	41 Nb				

The solar spectrum contains numerous metallic arc lines, the average ground-level excitation potentials of which are only ~1 eV. On the other hand, nonmetallic lines with high ionization potentials have also been encountered: He 20.9 eV, H 11.1 eV, N 10.0eV, O 9.3eV, C 7.5eV. S 6.5eV. However, these lines are few in number and, with the exception of hydrogen, are weak.

Some elements are represented in the solar spectrum by spark lines as well as arc lines, while others have spark lines only. Of these elements the most important (with respect to the number of lines) are Ti^+, V^+, Cr^+, Fe^+, Zr^+, La^+, Ce^+, Nd^+, Sc^+, Y^+, Sm^+. Table 3 lists the elements which, according to the Revised Rowland, have over 100 lines in the solar spectrum (neutral and ionized states combined).

Further work on the identification of solar lines, in conjunction with studies of laboratory spectra of the elements, will apparently lead to the discovery of still more elements in the solar atmosphere. The weakness (or even the total absence in the solar spectrum) of the lines of these elements may be attributed either to their low abundance (Ho, Re, Tl, Bi, Po, At, Rn, Ra, Ac, Pa, U) or to their high excitation potentials (Ne, A, Kr, X, Cl, Br, I, Se, Te, Cs, Hg).

TABLE 3

Element	Fe	Ti	Cr	Co	Ni	V	Mn	Zr	Ce	Sc	Ca	Y	Nd
No. of lines	3288	1085	1028	785	628	618	458	371	249	150	127	120	107

In addition to atomic lines, the solar spectrum (particularly sunspot spectra) includes molecular bands, of which the following have been identified: CN, C_2, CH, NH, OH, CaH, MgH, AlH, SiH, H_2, SiF, BO, AlO, TiO, FeO, and ZrO. The identification of SiN, CP, NaH, and PH is less certain.

Many lines and bands in the solar spectrum are telluric in origin. This is particularly the case in the red and infrared regions, where telluric lines predominate over solar lines. Telluric lines were first described by Brewster in 1890 and have since served as objects of study for many scientists. They differ from lines of solar origin in that their numbers and intensities vary considerably depending on the elevation of the sun above the horizon and on the state of the earth's atmosphere. This factor, and also the absence of Doppler shift due to solar rotation in the telluric lines present in the solar-limb spectrum, enable us to distinguish quite easily between lines of solar and terrestrial origin in the solar spectrum.

The most pronounced telluric lines are found to occur in groups which have a certain regularity of structure. Some of the more well-known line groups are the following: group α in the $\lambda\lambda$ 6272 —6335 region, including 38 oxygen lines; group B between λ 6868 and λ 6961, including 27 pairs of oxygen lines; group A between λ 7594 and λ 7699, including 27 pairs of oxygen lines; a group in the $\lambda\lambda$ 6929 — 7028 region, including 45 H_2O lines; and a group between λ 5862 and λ 6030, including 243 H_2O lines. Ozone lines occur in the ultraviolet spectrum, between $\lambda\lambda$ 2900 — 3150. On the whole, the telluric lines are mostly from O_2, N_2, O_3, CO_2, and H_2O. The presence of CO, N_2O, CH, OH, and some isotopic compounds of C^{12}, O^{16}, O^{18}, C^{13}, and O_2^{16} has also been demonstrated.

In recent years considerable progress has been made in the study of the infrared region of the solar spectrum. However, a huge amount of work still remains to be done in identifying the infrared lines recorded.

§9. Equivalent Fraunhofer-line widths. Growth curves. Studies of the profiles (see Vol. I, §179) of Fraunhofer lines give the most complete and thorough information on the physical state and structure of the solar atmosphere (photosphere). However, such studies are limited by the fact that the spectrographs used for measuring the true intensity distribution in the lines must meet very high specifications, mainly with respect to resolving power. Therefore, for spectroscopic solar research, and even more so for stellar spectroscopy, the total intensities or the

equivalent widths of absorption lines are usually used (loc. cit.), since these are independent of the resolving power of the spectrograph and thus can be determined from the observations with a high accuracy. It should be noted that visual estimates of line intensities on arbitrary scales (such as the Rowland scale) are at present of secondary significance. Equivalent widths constitute an effective means for obtaining various important physical parameters.

Data on the equivalent widths of Fraunhofer lines can be found in works such as that of Allen /3/ or else they can easily be obtained from the Utrecht photometric atlas of the solar spectrum. The equivalent width of a Fraunhofer line depends on the number of atoms participating in the formation of the line. The main task of the theory of Fraunhofer lines is to establish this dependence, which forms the basis of the study of the physical characteristics of stellar atmospheres and which, in particular, can be used to solve one of the most important problems in astrophysics, the quantitative chemical analysis of the stars and the sun.

Let us consider the simplest case, when the line originates due to the transmission of continuous radiation through a homogeneous absorbing layer, in which no re-emission takes place. We may write the following expression for the line profile:

$$I_\lambda = I_\lambda^0 e^{-k(\lambda)Rl}, \tag{1}$$

and by definition the equivalent line width is

$$W = \int_0^\infty \left(1 - \frac{I_\lambda}{I_\lambda^0}\right) d\lambda = \int_0^\infty (1 - e^{-k(\lambda)Rl})\, d\lambda, \tag{2}$$

where I_λ^0 and I_λ are the intensities of the incoming and outgoing radiation, respectively (for a given wavelength λ), l is the layer thickness in centimeters, R is the number of classical oscillators per cm^3, and k is the coefficient of absorption for a single oscillator. According to quantum theory $R = nf$, where n is the number of absorbing atoms per cm^3 and f is the oscillator strength corresponding to the given atomic transition. * It is clear from (2) that for very weak lines (optical layer thickness of $kRl \ll 1$) the equivalent width may be assumed to be proportional to the number of absorbing atoms, but for intermediate and large equivalent widths this dependence becomes more complicated and is determined by the variation of the absorption coefficient with wavelength (that is, it is determined by the line-broadening factors for the given case).

There can be various reasons for the broadening of lines in solar and stellar spectra: a Doppler effect due to the thermal motion of atoms and to gas motion; radiation damping; damping due to collisions with other atoms, ions, and electrons; a microscopic Stark effect (for hydrogen and helium lines); a Zeeman effect (for example, in sunspots); hyperfine structure and isotopic effects (for certain lines in the solar spectrum).

The thermal Doppler effect and radiation damping are the two line-broadening factors which have usually been studied the most thoroughly both in the laboratory and theoretically. In Volume I (§8) formulas were given for the absorption coefficient, taking into account these two factors (both

* The values of f for lines of astrophysical interest, and also a list of the relevant references, can be found, for example, in /4/.

separately and acting together). When both factors are present, the absorption coefficient has a complicated integral form:

$$k(v) = k_0 \frac{a}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-y}}{\left(\frac{a}{2}\right)^2 + (v-y)^2}, \tag{3}$$

where

$$k_0 = \frac{\sqrt{\pi}\, e^2}{mc^2} \frac{\lambda_0^2}{\Delta\lambda_D} \tag{4}$$

is the coefficient of absorption at the center of the λ_0 line, and where

$$a = \frac{\Delta\lambda_N}{\Delta\lambda_D} = \frac{\gamma_0}{\Delta\omega_\infty} \text{ and } v = \frac{\lambda_0 - \lambda}{\Delta\lambda_D} \tag{5}$$

are dimensionless auxiliary terms introduced for simplicity. Here, $\Delta\lambda_D$ (or $\Delta\omega_D$ in terms of angular frequencies) is the Doppler broadening, $\Delta\lambda_N$ = $= 1.18 \cdot 10^{-4}$ Å is the natural line width, and $\gamma_0 = 0.220/\lambda_0^2$ is the radiation-damping constant. * In the case of a mixed Doppler effect (for example, due to both thermal atomic motion and to turbulence), supplemented by damping of various kinds (for example, both radiation and collision damping), the absorption coefficient can be expressed using the same formula (3), but the Doppler constants must be combined by adding up the squares of their values and the damping constants must be simply added together. It should be noted that, according to Minnaert and Mulders, the Doppler broadening and the damping are sufficient to describe almost all the aforementioned effects, with the exception of Stark broadening. Various authors have tabulated the integral in (3), and also convenient series expansions have been derived enabling a determination of the absorption coefficient for various values of a and v in (5).

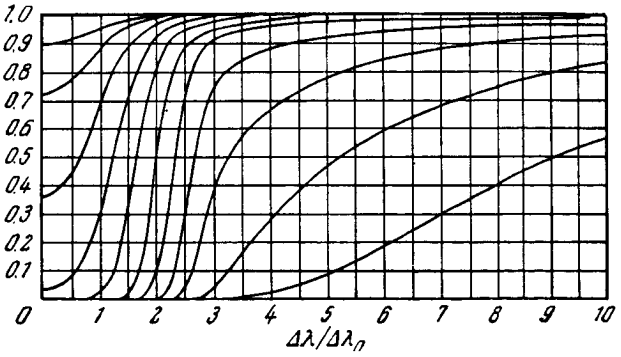


FIGURE 8

The broadening of Fraunhofer lines is due mostly to the combined influence of the Doppler effect and damping of various kinds. However, under stellar atmospheric conditions the Doppler broadening $\Delta\omega_D$ exceeds the damping width γ considerably, that is, $a \ll 1$. The absorption coefficient

* The notation used here is identical to that in Vol.I, §8.

(3) in this particular case, which is the only case of astrophysical significance, becomes

$$k(v) = k_0 \left[e^{-v^2} - \frac{a}{\sqrt{\pi}} \left(1 - 2ve^{-v^2} \int_0^v e^{z^2} dz \right) \right]. \qquad (6)$$

This expression shows (see Vol. I, §8) that near the center of the line (small v), the absorption coefficient is determined by the Doppler effect only ($k \sim e^{-v^2}$), whereas in the outer parts of the line (large v) it is determined only by the damping ($k \sim \frac{1}{v^2}$). Figure 8 illustrates the variation in line profile as the number of absorbing atoms increases. Initially the line has a Doppler profile; the depth of the line increases rapidly with Rl, but its half-width varies only gradually, especially when the line center becomes completely dark. Subsequently, extended wings appear due to damping, then the wings intensify, and finally the profile assumes a form which does not change when Rl is further increased. The total absorption W for a line varies in accordance with the profile. For $Rl \leqslant 10^{10}$ cm^{-2} the total absorption increases in proportion to Rl, after which the growth of W slows down, until at $Rl \approx 10^{15}$ or 10^{16} (when the wings appear) W again starts increasing at a higher rate. For $Rl > 10^{16}$ the total absorption (and also the halfwidth of a line) increases in proportion to $\sqrt{Rl\gamma}$.

Equation (2) cannot be integrated in its general form, but integration is possible in particular cases and the formulas obtained are convenient for calculating the equivalent widths. For instance, in the case of pure Doppler broadening ($a = 0$), the following series expansion is obtained:

$$\frac{W}{2\Delta\lambda_D} = \frac{\sqrt{\pi}}{2} X_0 \left(1 - \frac{X_0}{2!\sqrt{2}} + \frac{X_0^2}{3!\sqrt{3}} - \frac{X_0^3}{4!\sqrt{4}} + \cdots \right), \qquad (7)$$

which applies for $X_0 < 3$, where $X_0 = k_0 Nf$ is the optical thickness of the layer at the line center, with $N = nl$. This formula is also valid in the presence of damping, provided the values of Nf are sufficiently small (at least up to $W < \Delta\lambda_D$). Thus, for $a \ll 1$ we have

$$W = \sqrt{\pi}\, X_0 \Delta\lambda_D = \frac{2\pi e^2 l^2}{mc} Nf. \qquad (8)$$

On the other hand, for very large values of Nf, when the broadening is determined by damping only, the equivalent width may be expressed as

$$W = \sqrt{\frac{4\pi^2 e^2}{mc} Nf\gamma}. \qquad (9)$$

In cases when both line-broadening factors are significant, so that the absorption coefficient must be expressed in its general form (6), W may be determined by numerical integration.

The curve relating the equivalent widths W with the number of absorbing atoms $N \times f$ (both variables are generally plotted on a logarithmic scale) is called the curve of growth. Figure 9 shows growth curves for various values of parameter a, according to Held. On the ordinate axis is plotted $\lg \frac{W_\lambda}{2\Delta\lambda_D}$ and on the abscissa axis is plotted $\lg \frac{Rl}{\Delta\omega_D} = \lg \left(\frac{\lambda_0^2}{2\pi c} \frac{Nf}{\Delta\lambda_D} \right)$. On curves

obtained from astrophysical observations ($a \ll 1$) three regions may be clearly distinguished, namely the linear region, the "saturation" region, and the wing region; in all of these regions the equivalent width depends only on Rl and is mainly determined by the Doppler effect or by damping. *

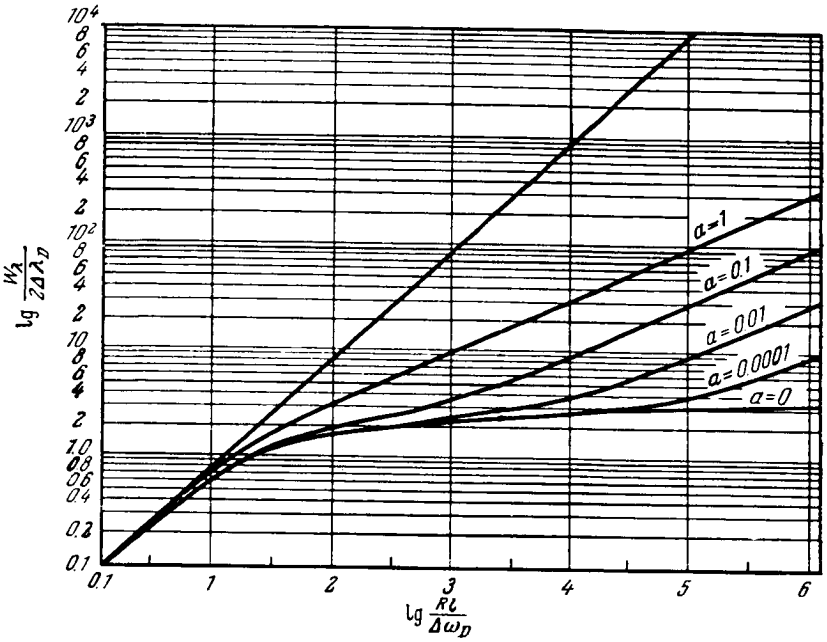


FIGURE 9

The application of growth curves to solar and stellar research presupposes the possibility of plotting the corresponding curves directly from observations as well. This can be done using multiplets whose components have known relative oscillator strengths f_{ik}. We choose a multiplet whose components have a common ground level, and thus an almost equal number N_i of absorbing atoms, and then we plot $\lg W$ as a function of $\lg f_{ik}$ to obtain a section of the growth curve. The various curve sections obtained by this method from different multiplets are next joined to form a smooth curve by means of translation along the abscissa axis. Since a growth curve depends on wavelength, as shown by formulas (7), (8), and (9), the quantity $\lg(\frac{W}{\lambda})$ or $\lg(\frac{W}{\lambda}\frac{c}{\xi_0}) - \lg\frac{W}{\Delta\lambda_D}$ (where ξ_0 is the most probable velocity) is substituted for $\lg W$ in order to eliminate this dependence when matching multiplets from different spectral regions to form a single curve. ** Although, strictly speaking, only lines with a common ground level should be used, since only then is it certain that the lines will originate in the same solar layer, in practice the solar atmosphere as a whole is often considered and an average growth curve is plotted combining multiplets which originate in different layers.

* For a comprehensive theory of growth curves, see /6, 7/.

** In this case λ also enters into (9), but in the radicand.

If an "observed" growth curve of $\lg\left(\frac{W}{\lambda}\right)$ as a function of $\lg f$, drawn to scale on tracing paper, is moved across the graph in Figure 9, then we can find the theoretical growth curve which agrees best with the observed data. This establishes the value of parameter a. The velocity ξ_0 is then obtained from a comparison of the ordinates of the two curves. Next, given a and $\Delta\lambda_D$, we can derive the damping constant γ. Once the observed and theoretical growth curves have been matched, we can use the observed value of $\lg\left(\frac{W}{\lambda}\right)$ to find the corresponding value of $\lg Nf$ for any line, and thus if the absolute value of f is known we can compute the number of atoms N.

Once we have determined the relative populations of the various atomic levels from the growth curve, we can verify the Boltzmann distribution and calculate the corresponding excitation temperature. In order to do this, however, we may also apply a somewhat different procedure for constructing the growth curve /6/. First, to construct the growth curve we measure along the abscissa axis the quantity

$$\lg X_f = \lg gf\lambda_0 - \frac{5040}{T}E, \tag{10}$$

where gf is the observed laboratory line intensity (g being the statistical weight), E is the excitation potential of the multiplet ground level, and T is the excitation temperatures for the given atoms. The temperature and all other physical quantities are obtained from the growth curve by successive approximations.

Many authors have constructed growth curves for the sun. Until recently, only average growth curves were plotted, particularly since the initial data (the f-numbers and equivalent widths) for individual elements were few and inaccurate. Valuable contributions to growth-curve construction were made by the laboratory work of A. King and R. King (who determined the relative f-numbers for many lines of Fe, Ti, V, and other elements), by the studies of several physicists of Rozhdestvenskii's school, and by many other scientists.

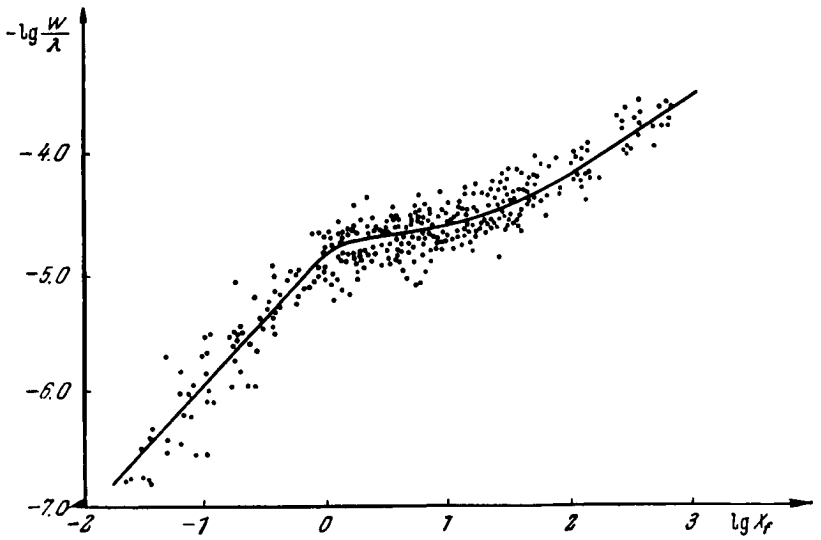


FIGURE 10

Figure 10 gives the most complete and accurate growth curve for Fe I, constructed by Mitrofanova (Pulkovo Observatory) in 1952. It was plotted using 178 multiplets comprising 650 spectral lines. The laboratory intensities *gf* for the lines covering the upper and middle parts of the growth curve were remeasured by Vyazanitsyn, while theoretical intensities were used for the lower part of the curve, which corresponds to lines which are excited with difficulty. The equivalent widths used were the averages of three independent values: those obtained from special observations, those taken from the Utrecht photometric atlas of the solar spectrum, and those borrowed from Allen's tables.

TABLE 4

Author	FeI	TiI
Menzel, Baker	4150° ± 50°	4350° ± 200°
Goldberg		4400
Wright	4750 ± 150	4575 ± 125
Goldberg and Pierce	4770	4640 ± 290
Gottschalk	4800 ± 200	
Mel'nikov	4580 ± 100	
Mitrofanova	4535 ± 100	

As observed previously, one of the most important physical characteristics which can be obtained from the growth curve is the excitation temperature of the reversing layer of the sun. The excitation temperatures found from the lines of different elements vary over a wide range. For example, from the Cr I lines Sandage and Hill (1951) obtained a value of T = 3790°. On the other hand, the V I lines gave 5400° (King and Wright, 1947) and 5110° (Sandage, 1950). The results of several determinations of the excitation temperatures from the Fe I and Ti I lines are given in Table 4. Apparently, the observed differences in T for the different elements actually exist. The fact that these values of T are generally lower than the effective solar temperature (5710°), although sometimes they are close to it, still has not been explained convincingly.

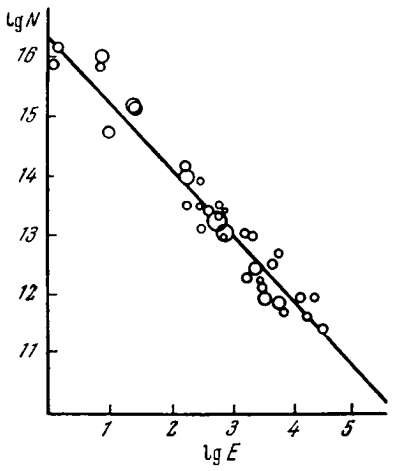


FIGURE 11

In order to verify to what extent the assumption of a Boltzmann distribution of atoms according to energy states is satisfied, the assumption upon which the excitation-temperature distribution was based, we should compare the values of lg N (the number of atoms in a given state per cm^2 along the line of sight) and E (the excitation potential). The corresponding graph for the FeI atoms, plotted by Mitrofanova (Figure 11), indicates a linear

dependence between the two quantities, and thus confirms the applicability of Boltzmann's formula to the reversing layer for E from 0 to 5 eV.

The halfwidths of the weak Fraunhofer lines, as noted by several authors, exceed the Doppler halfwidth considerably, due to thermal motion of atoms. This led to the conclusion that turbulent motion exists in the reversing layer. Different investigators (Redman, Allen, ten Bruggencate, and van de Hulst) found velocities of 1.0 to 1.8 km/sec (radial component). On the basis of empirical growth curves, some authors obtained velocities of similar magnitude: 0.9 (Wright), 0.6 (Rubinstein), 1.2 (Claas), 0.0 [1.0 ?] (Mitrofanova), and 1.5 km/sec (Hubenet). Thus, the turbulent velocities in the reversing layer vary between about 1.0 and 1.5 km/sec.

The damping constant has been determined from the growth curves by various authors. The observed damping constant γ for the sun is in most cases substantially higher than the radiation damping constant γ_0. For example, Minnaert and Mulders, and also Allen, found a value of $\gamma/\gamma_0 = 10$, Wright found 15, Voigt found 100, Bell found values from 14 to 300, etc. These results indicate that collision damping is of considerable importance in the reversing layer of the sun, in addition to the radiation damping there.

§10. The quantitative composition of the solar atmosphere. The chemical composition of the solar atmosphere was first investigated quantitatively by Russell in 1929, on the basis of the Revised Rowland, the only source then published. Russell's method essentially amounts to a calibration of Rowland's estimates of intensity for the solar lines in multiplets, in terms of the relative "numbers of atoms."* Then, if we know the average temperature and average electronic pressure in the solar atmosphere (Russell gives $T = 5040°$ and $p_e = 3.1$ bars), it is possible to apply the formulas of Boltzmann and Saha to find the total number of atoms of the element in question. Table 5 gives the results of calculating the numbers of atoms of various elements per cm^2 of solar surface in Russell units. In order to convert to absolute values of N, the normalizing factor $N_0 = Nf$ must be introduced for the line of Rowland intensity 0; this factor was derived by Russell by comparing his results with the Nf values obtained by Unsöld from observations of line profiles assuming pure radiation damping, and the value obtained was $\lg N_0 = 12.67$.** Despite the manifest simplification and shortcomings of this method (for instance, the assumption of a single calibration curve for all lines regardless of intensity, etc.), Russell's results have not in general been modified essentially by later studies. For some of the heavier elements, Russell's data are still the only source of information available.

Subsequently, however, considerable progress was made in the field of the chemical analysis of the sun and stars, on the basis of the development of Fraunhofer-line theory and due to the use of exact measurements of the line intensities instead of visual estimates on an arbitrary scale. At present, two basic methods are used to determine the chemical composition of the sun. The first method amounts to an application of the general theory of growth curves, which was illustrated in the preceding section. The equivalent width and the growth curve are used to determine the population of a certain atomic level, and then the Boltzmann and Saha formulas are

* One unit on the Russell scale corresponded to the number of atoms reproducing a line of Rowland intensity 0.

** According to Unsöld, $\lg N_0 = 11.8$ (see Reference /7/). The possible error in the Russell logarithms for the contents of the elements is estimated by Unsöld as +1.

used to compute the relative populations of all the other excitation and ionization levels (the average temperature and average electron pressure are assumed known). As a result, it is possible to determine the total content of atoms of the given element in the solar atmosphere. Because of the imperfection of the present theory of growth curves, this method does not always ensure satisfactory accuracy and reliability of the results, but its relative simplicity and certain other features make it the most applicable one. The main defect of this method is that it proceeds from an extremely simplified model of the solar atmosphere, which is visualized as a scattering reversing layer above the photosphere, the optical thickness of the layer being assumed constant for all lines.

TABLE 5

Element	lg N	Element	lg N	Element	lg N	Element	lg N
H	11.5:	Ti	5.2	Y	2.6	Ce	2.4
Li	2.0:	V	5.0	Zr	2.5	Pr	0.6:
Be	1.8	Cr	5.7	Nb	1.0:	Nd	2.0
C	7.4:	Mn	5.9	Mo	1.4	Sm	1.5
N	7.6?	Fe	7.2	Ru	1.7	Eu	1.4:
O	9.0:	Co	5.6	Rh	0.5	Gd	1.1:
Na	7.2	Ni	6.0	Pd	1.1	Dy	1.6:
Mg	7.8	Cu	5.0	Ag	1.0	Er	0.1:
Al	6.4	Zn	4.9	Cd	2.2:	Hf	0.4
Si	7.3	Ga	2.0	In	0.0:	W	0.2
S	5.7:	Ge	3.0	Sn	1.2	Ir	0.2?
K	6.8:	As	0.6?	Sb	0.8	Pt	1.6
Ca	6.7	Rb	1.7:	Ba	3.3	Te	1.4?
Sc	3.6	Sr	3.3	La	1.8	Pb	1.2

TABLE 6

Element	Menzel and co-workers	Unsöld	Claas	Hunaerts	Element	Menzel and co-workers	Unsöld	Claas	Hunaerts
H	24.40	24.40	24.40	24.40	V	16.49	16.46	—	—
He	23.70	23.70	—	—	Cr	17.27	17.98	—	—
Li	—	—	13.48	—	Mn	17.49	17.86	—	—
C	19.96	20.69	—	21.46	Fe	19.39	20.12	20.02	—
N	20.49	21.01	—	21.42	Co	17.09	17.43	—	—
O	20.96	21.13	21.05	21.63	Ni	18.79	18.36	—	—
Na	18.96	18.68	18.73	—	Cu	16.79	16.63	17.20	—
Mg	20.79	19,91	19.97	—	Zn	17.97	17.18	16.92	—
Al	18.79	18.73	18.57	—	Sr	—	15.75	15.28	—
Si	20.27	19.69	19.52	—	Y	—	15.61	—	—
S	19.97	19.32	—	—	Zr	—	14.77	—	—
K	17.49	17.60	17.41	—	Mo	—	14.18	—	—
Ca	18.97	18.63	18.86	—	Ba	—	15.35	14.78	—
Sc	—	15.73	—	—	Pb	—	15.0	—	—
Ti	16.97	17.36	—	—	Hg	—	—	15.38	—

The second method for determining the chemical composition is more exact. It is based on the study of line profiles and uses a model which takes into account both the scattering and the actual absorption. In this method a certain initial chemical composition is assumed and the absorption parameters are computed for various levels in the photosphere; integration of the transport equation for several values of the relative abundance of the element then indicates the line profile which best agrees with the observed

profile and thus gives the abundance of the element. Once the abundances of all the elements are determined, we may refine our picture of the structure of the photosphere, etc. A definite advantage of this method is that it requires the use only of weak and and very weak lines (in addition to other factors, these lines are independent of the impact-damping constants, which are inaccurately known and vary with depth). The application of this method does not rule out the construction of growth curves as well, but they must have a more rigorous theoretical basis.

A great number of investigations of the chemical composition of the solar atmosphere have been made using these methods in their different alternative forms. Table 6 gives the results of some recent studies /8/. The values of $\lg N$ are listed, where N is the number of atoms of the given element per cm^2 of solar surface; for hydrogen it was assumed in all cases that $\lg N = 24.40$. The most comprehensive analysis was that of Unsöld, who applied the growth-curve method. Claas mainly used the method of weak lines. Hunaerts derived the abundances of the elements from an examination of molecular spectra. The satisfactory agreement of the abundances obtained by the different methods verifies their actual worth. However, the error in the relative abundances ($\lg N$) was still as high as 0.5 to 0.7 or even more. For the hydrogen content different authors give values ranging from 10^{24} to 5.10^{24} atoms/cm^2.

The inaccuracy in the determination of the element abundances is in some cases (for the lines of H, He, C, N, O, Si, etc., and especially for ionized elements) due to the uncertainty in the Boltzmann correction for transitions from high energy levels (> 5 eV). On the other hand, the solar spectrum contains no He lines at all except for λ10,830. All these cases require special analysis and verification. For example, from the molecular bands of CH, OH, CN, NH, and C_2 in the solar spectrum Hunaerts (1950) determined the abundances of H, C, N, and O. The abundance of oxygen was also determined according to the weak forbidden lines, which are absorbed by low atomic levels and which were recently discovered in the solar spectrum (Bowen, 1948; Cabannes and Dufay, 1948). According to Gauzit, who identified two weak infrared lines of N I, its abundance is almost equal to the abundance of O I. The helium content of the sun can be identified with the helium existing in the chromosphere and the prominences, although the estimate will remain uncertain as long as the mechanism of the helium emission in these formations remains insufficiently clear.

TABLE 7

Elements	Number of atoms (in %)
H	83.224
He	16.645
C, N, O	0.121
Mg, Si, Fe	0.0083
Other metals	0.0013

For illustrative purposes it is convenient to break down the composition of the solar atmosphere according to groups of elements, as is done in

Table 7 /8/. This table illustrates several interesting features of the chemical composition of the solar atmosphere. The most remarkable fact is the predominant abundances of the two lightest elements, H and He. Hydrogen atoms constitute 83.2% of all the atoms and are 5 times more abundant than helium atoms, which constitute 16.6%. Thus, the sum total of the atoms of hydrogen and helium together makes up 99.8% of all the solar atoms. The most abundant elements after H and He are C, N, and O, so that the solar atmosphere can actually be regarded as composed of these five elements. Elements with atomic numbers greater than 10 occur as "impurities." Of the metals, Mg and Fe (which are common on the earth) are the most abundant. The total number of atoms of these metals constitutes 86% of all the metal atoms in the solar atmosphere. Hydrogen atoms are 8 or 9 thousand times more abundant than the sum for all the metals. In general, the abundance ratio of hydrogen atoms to metal atoms represents an important astrophysical constant, one which has been determined by many authors. The best results are apparently those of Strömgren and of Claas, who give $\lg \left(\frac{N_H}{N_m}\right) = 3.9$ and 3.95, respectively.

TABLE 8

Compound	OH	CH	NH	CN	C_2
lg N	17.52	17.31	16.98	16.32	14.96

The most abundant chemical compounds in the solar atmosphere are OH, CH, NH, CN, and C_2. The numbers of molecules of these compounds, according to Hunaerts (1948), are given in Table 8. It is evident that the total content of molecular compounds in the solar atmosphere is even far less than the content of metals.

Finally, let us consider the results of some recent determinations of the content of deuterium (D or H^2) in the solar atmosphere. The deuterium line is wide and is observed in the form of a depression of about 1 or 2% on the wing of the Balmer line. A new determination of the deuterium abundance was made by de Jager, who found $D/H = 1.45 \cdot 10^{-4}$, which is fairly close to the terrestrial abundance. Shklovskii obtained $D/H \leqslant 10^{-6}$ according to the flash spectrum. From photoelectric measurements of the profiles of the H_α and H_β lines, Severnyi found a solar deuterium abundance which was several times greater than the terrestrial abundance.

REFERENCES FOR CHAPTER II

1. MINNAERT, M., G.F.W.MULDERS, and J.HOUTGAST. Photometric Atlas of the Solar Spectrum. — Amsterdam, 1940.
2. KUIPER, G. (ed.). The Solar System. Vol.I, The Sun, Ch.IV (Moore, C.E. The Identification of Solar Lines). — Russian Translation from English, Edited by V.A.Krat. 1957.
3. ALLEN, G.W. Astrophysical Quantities. — Univ. of London. 1955.

4. AMBARTSUMYAN, V.A. et al. Teoreticheskaya astrofizika (Theoretical Astrophysics). — Gostekhizdat. 1952. [English Translation. Pergamon.]
5. UNSÖLD, A. Physics of Stellar Atmospheres. — Russian Translation from German, Edited by E.R. Mustel', Izdatel'stvo inostrannoi literatury. 1949.
6. MITROFANOVA, L.A.—Izvestiya Glavnoi astronomicheskoi observatorii, No.149: 81. 1952.
7. Transactions of the International Astronomical Union, Vol. VII. 1950.
8. MITCHELL, S.A. Eclipses of the Sun. — 1953.

V.A.KRAT

N67-15487

Chapter III

THE STRUCTURE OF THE PHOTOSPHERE: GRANULATION, SUNSPOTS, AND FACULAE

§11. The general appearance of the solar surface. Granulation. The surface of the sun (when examined on a screen or photograph) consists of individual tiny bright spots projected onto a darker background. Since these tiny spots sometimes resemble rice grains (granules), the nonuniformity of the brightness distribution, between the tiny bright spots and the background, is known as granulation. In addition, sunspots appear, develop, and eventually disappear on the solar disk; these sometimes attain prodigious size and may even be visible to the naked eye. A sunspot consists of two basic parts: the central, darkest part, called the umbra, and the lighter surrounding fringe, called the penumbra. Penumbrae are observed for large spots only. Sunspots generally form into groups, sometimes comprising several tens of spots of various sizes. Near spot groups, and in some case independently of them as well, strings of very bright granules, called faculae, are observed. Faculae, like sunspots, are transient formations on the solar surface. All these formations occur in the layer of the sun known as the photosphere.

The photosphere is the immediate source of almost all the solar radiation in the visible spectrum. The absorption and emission of light over a continuous frequency range occurs in the photosphere, and since the photospheric gas is partly transparent, radiation reaches us from various depths in the photosphere. However, no direct luminous flux is received from depths greater than 400 km, and so the thickness of the photosphere has sometimes been set at 400 km. This conclusion is a fairly arbitrary one, though, since this value was not obtained from observations but rather from theoretical computations. An intensive emission and absorption of light (mostly by scattering) in the main range of the Fraunhofer lines takes place in the photosphere. The modern conception of the photosphere can therefore be identified with the older idea of a reversing layer, which was previously thought to be a gas layer located above the photosphere and producing just the line spectrum, the photosphere being the source of the continuous spectrum.

At a depth of about 400 km below the upper boundary of the photosphere, practically in the lowest photospheric layers, the convective zone begins. Since the opacity of the gas rises rapidly with the increasing gas density at greater depths, the temperature gradient of a gas in radiative equilibrium should become more adiabatic at sufficient depths. Under such conditions, convection sets in automatically, and the energy transfer occurs mainly by convection. In this process the increase in the degree of ionization of hydrogen is of great significance, since it reduces the ratio of the specific

heats of the gas at constant pressure and volume (the adiabatic index), and consequently reduces the adiabatic temperature gradient as well. At the center of the disk the convective zone starts at an optical depth of $\tau = 2$. Therefore, about 13% of the radiation in this case reaches us directly from the convective zone, and the result is that the photosphere appears to be translucent to it.

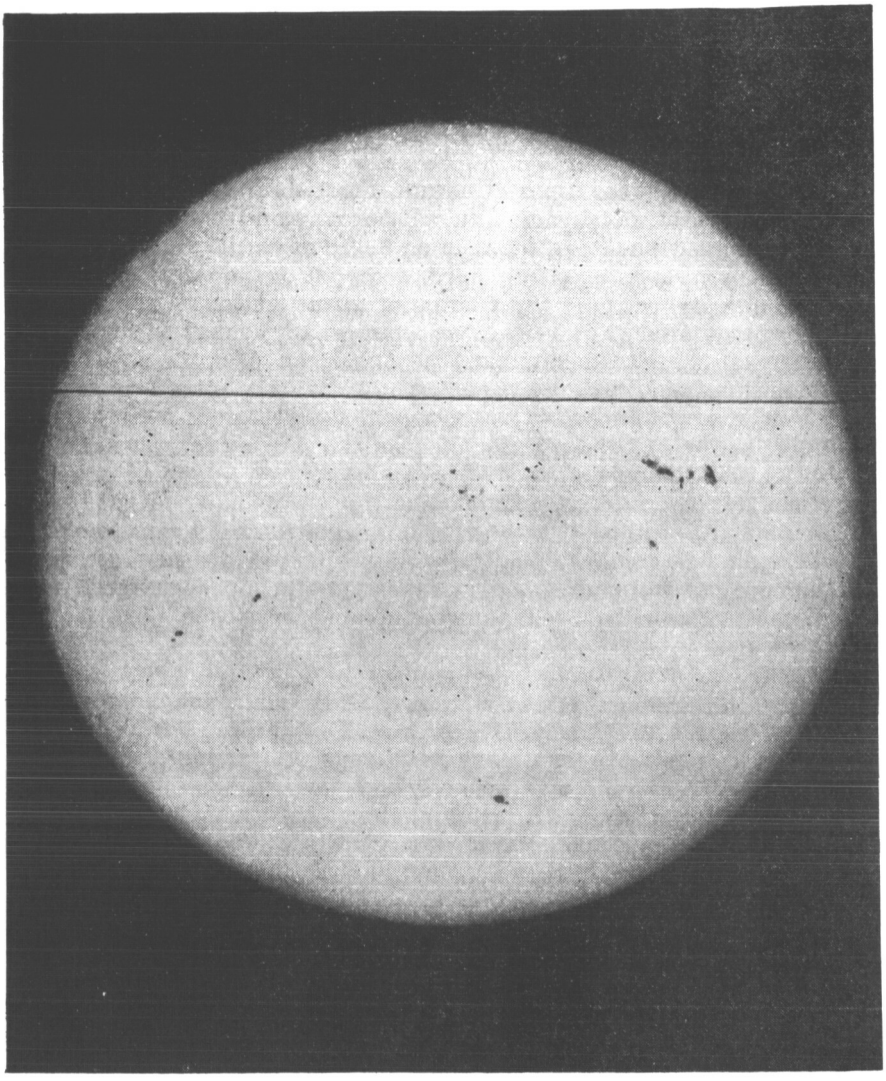


FIGURE 12

Figure 12 shows a photograph of the sun taken at Pulkovo Observatory with a horizontal solar telescope. Individual spots and faculae are visible.

Transient solar formations, sunspots and faculae included, are known to occur in greatest number at more or less regular intervals, and the

average yearly number of observed sunspots and faculae varies with a period of about 11 years. During the epochs of maximum, spots and faculae are observed in abundance, whereas during epochs of minimum, the solar disk is often almost entirely devoid of these formations.

The study of the granulation is quite difficult, since the angular dimensions of the individual granules are exceedingly small. Observations can be made only using high-quality optical instruments and only when the solar image is exceptionally quiet and clear, which is very rarely the case. Air turbulence causes the image to fluctuate and to become blurred, which greatly complicates the observations. In addition, streams of heated air, mainly in the surface layer near the observation point but sometimes in the instrument itself, can blur many of the features of a solar photograph. Finally, certain parts of the photograph may appear slightly displaced relative to other parts, or else sometimes distinct and blurred parts of a photograph can alternate to form a more or less regular grid (previously such patterns were thought to be actual solar phenomena (Janssen)).

Detailed photographic studies of the granulation were initiated by Janssen (1877–1894) at Meudon /1/, by Ganskii /2/ (1905–1907) at Pulkovo, and by Chevalier (1906–1912) at the Zô-Sè Observatory in China. The most complete and reliable information on solar granulation has been that obtained in recent years with the horizontal solar telescope at Pulkovo and with a special camera carried by a stratosphere balloon to a height of 32 km (Schwarzschild, USA).

The smallest granules do not exceed 100 to 150 km in diameter. Because of the tendency of small granules to collect into groups (sometimes including a hundred or more granules which almost merge together), the so-called medium-sized granules (400 km = 0".7 in diameter) can actually be regarded as groups of 3 or 4 small granules. The lifetimes of these granules vary for the different parts of the solar surface and for different years; the average lifetime ranges from 1 to 2 minutes.

Most of the granules, as shown by the Pulkovo observations and in particular by the excellent photographs of Schwarzschild and his co-workers, are about 1".5 (1000 km) in diameter. These are the so-called "second-of-arc" granules, which can also be observed without difficulty under terrestrial conditions. Such granules have an average lifetime of about 5 or 6 minutes (according to Schwarzschild). The observations of Kerimbekov, made at Pulkovo somewhat earlier with a motion-picture camera, gave almost the same results. Thus, there is reason to believe that, the larger the granule or group of granules, the longer will be the average lifetime.

The granulation is most easily discerned in the central parts of the solar disk (Figure 13, photograph taken with Pulkovo horizontal solar telescope). At the limb the granules undergo foreshortening, become narrow, and blend together. On the basis of his observations, Chevalier concluded that the granulation has the very same appearance during both maximum and minimum solar activity. However, the Pulkovo observations for the period from 1951 to 1957 /3/ show that Chevalier's conclusion was somewhat premature, since the nonuniformity of the granule distribution over the solar disk varies from year to year. The contrast between the granules and the darker background between them is very slight. The Pulkovo observations give an average contrast of 5 to 6% in the photographic part of the spectrum.

Photoelectric observations at Pulkovo have revealed an interesting feature of the granulation. The granules are observed as fluctuations superimposed onto large-scale background fluctuations. Consequently, the impression is formed that certain large waves in the photosphere become dissipated, decaying to form smaller waves with lower amplitudes and wavelengths. The total amplitude of a brightness fluctuation may be as much as 10 to 13%.

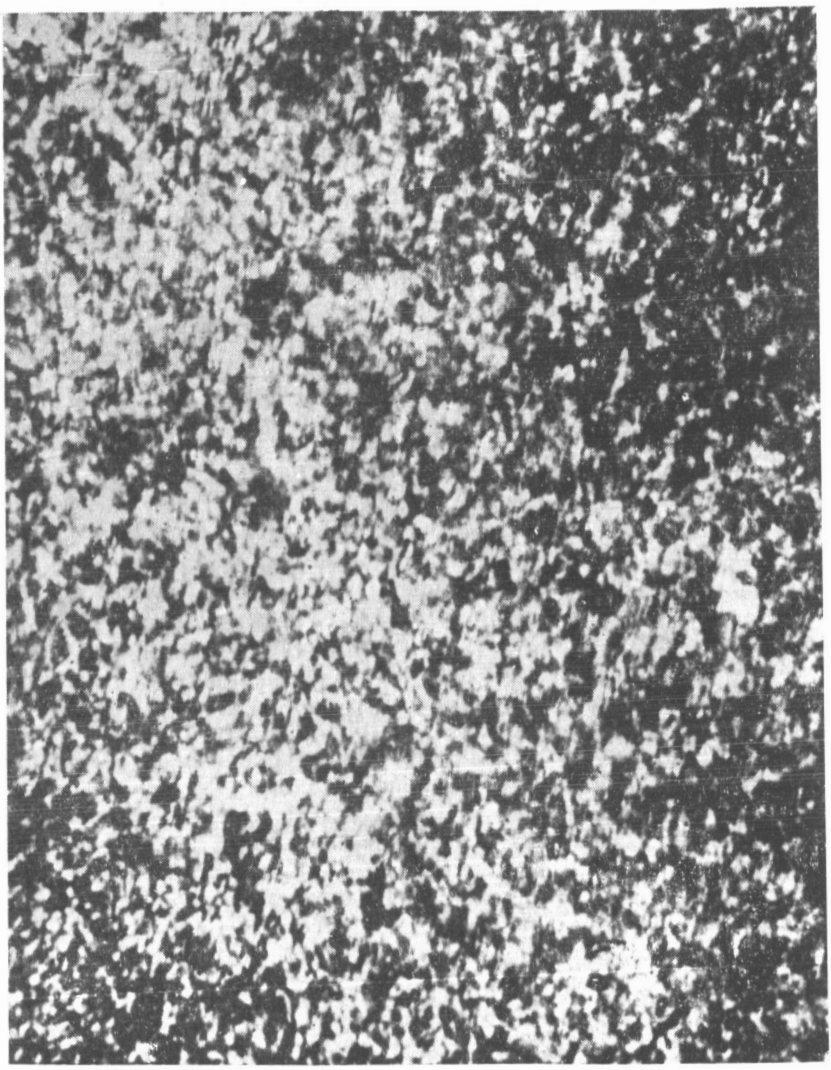


FIGURE 13

The low lifetimes of the granules, together with the absence of stable small features on the solar surface from which the granule coordinates could be reckoned, render impossible a direct determination of the horizontal motion of the granules or their motion in an image plane over the solar

surface. The granule velocities determined previously by Ganskii were shown to be illusory by Macris, on the basis of a study of a series of granulation photographs obtained by Lyot at Meudon. This conclusion has been verified by recent results at the Pulkovo Observatory. The order of magnitude of the granule velocities in the photosphere can be determined only by measuring the radial [line-of-sight] velocities. According to Richardson and Schwarzschild, who obtained a good spectrogram of the granulation field, and according to the many granulation spectrograms obtained by Plaskett, random motions with velocities of some tenths of a kilometer per second prevail above the granules in the reversing layer. However, this conclusion pertains not to individual granules but to granule groups 2000 to 4500 km in size.

It is quite difficult to obtain good granule spectrograms, since photography of the solar spectrum onto a fine-grain emulsion, using a fairly high resolving power and a high dispersion, requires comparatively long exposure times, of the order of several tenths of a second (direct photographs of the photosphere are taken with exposures of about 0.001 sec). During such long exposures, a strong blurring of the solar disk due to atmospheric vibrations may be produced. Thus, granule spectra can be obtained only under exceptionally favorable conditions. In 1960 at Pulkovo, granulation spectrograms were obtained which showed the spectra of the "second-of-arc" granules. It was established that the radial velocities for these granules are virtually the same as those for granule groups. Only for the sodium D_1 and D_2 lines, whose cores originate in the uppermost photospheric layer, were the radial velocities much higher. On the average these velocities were 1.7 km/sec. It is interesting that there is no distinct dependence between the brightness fluctuations and the radial velocities. An increase in brightness may correspond to either ascent or descent. Only a slight correlation exists to indicate that an increase in brightness often accompanies an upward motion of the gas. All this can be explained provided we remember that the brightness fluctuations are actually produced in the deepest layers of the photosphere (those of higher temperature), whereas the absorption lines originate in the intermediate and upper parts of the photosphere. The distance between these layers amounts to 200 or 300 km. If the gas motion in the photosphere is considered to be a propagation of acoustic waves (or, more precisely, of gravitational-acoustic waves) generated by convective currents below the photosphere, then the fluctuations observed within the photosphere will naturally have a different phase. The fluctuation amplitude should increase with a decrease in the gas density, reaching a maximum value at the boundary of the photosphere.

In certain regions of the photosphere, mainly near spot groups and faculae, large-scale motions are observed which cover areas of the order of 10^9 or 10^{10} km^2. In some places the gases ascend, while in others they descend, and the velocities in individual instances may reach two or three kilometers per second. The presence of such motions makes it impossible to determine the rotation velocity of the sun accurately from measurements of the radial velocities (Chapter I), and large systematic errors may creep in.

The strings of faculae and the spot penumbrae also consist of individual granules. The facular granules are on the average brighter than the ordinary photospheric granules, although the two kinds of granules do not differ as far as lifetimes and sizes are concerned. The background

against which strings of faculae appear is generally brighter by a few percent than the intergranular background of the undisturbed photosphere. The strings of faculae and the large facular granules are usually exceptionally stable and may remain unmodified for an hour or more. On the other hand, they can also form and disappear almost instantaneously, within the course of a minute. While they are easily visible near the solar limb, where they sometimes appear as prominences, facular granules are hardly discernible near the center of the disk. Near the limb the contrast between the brightest facular granules (according to the data of the latest Pulkovo photographic observations) and the surrounding background is from 30 to 40%, while near the center the contrast with the photospheric background is only a few percent higher than the average contrast of the ordinary granules. This situation can naturally be explained by the fact that a facular granule is situated over a photospheric region whose temperature is lower than the average for the photosphere. In other words, the temperature gradient in the facula is lower than that in the undisturbed photosphere. Because of this, T is greater in the upper layer and less in the lower layer than the value in the surrounding gas.

According to Eddington's approximation, the temperature T of the photosphere is related to the optical depth τ by the following formula /4/ (gray matter):

$$T^4 = T_0^4\left(1+\frac{3}{2}\tau\right), \tag{1}$$

where T_0 is the temperature at the boundary of the photosphere. Such a variation of T with τ leads to the following brightness distribution over the solar disk:

$$B(\vartheta) = B\left(\frac{\pi}{2}\right)\left(1+\frac{3}{2}\cos\vartheta\right), \tag{2}$$

where $B(\vartheta)$ is the brightness at a distance ϑ from the disk center. The coefficient preceding cos ϑ represents the gradient of the variation of $\left(\frac{T}{T_0}\right)^4$ with τ. The brightness distribution in a facula is different:

$$B_f(\vartheta) = B_f\left(\frac{\pi}{2}\right)(1+u\cos\vartheta), \tag{3}$$

where $u < 1.5$ at the disk center and $B(0) = B_f(0)$. Formula (1) for a facula becomes

$$T_f^4 = T_{0,f}^4(1+u\cos\vartheta). \tag{4}$$

When (3) is divided by (2), we obtain

$$\frac{B_f(\vartheta)}{B(\vartheta)} = \frac{1+u\cos\vartheta}{1+1.5\cos\vartheta}\,\frac{B_f\left(\frac{\pi}{2}\right)}{B\left(\frac{\pi}{2}\right)}. \tag{5}$$

Since the radiation at the solar limb comes essentially from the surface of the photosphere, therefore

$$\frac{B_f\left(\frac{\pi}{2}\right)}{B\left(\frac{\pi}{2}\right)} = \left(\frac{T_{0,f}}{T_0}\right)^4. \tag{6}$$

According to the Pulkovo observations, the integrated-light contrast between the brightest facular granules near the limb and the photosphere is 1.69 (for $\overline{\cos\theta} = 0.231$). If we insert this value into (5) and take into account that

$$\frac{B_f\left(\frac{\pi}{2}\right)}{B\left(\frac{\pi}{2}\right)}\frac{1+u}{2.5}=1, \tag{7}$$

then we obtain

$$u=0.13 \text{ and } \left(\frac{T_{0,f}}{T_0}\right)^4=2.2.$$

Thus, setting $T_0 = 4950°$, we have

$$T_{0,f}=\sqrt[4]{2.2}\cdot 4950°\approx 6000°.$$

For $\tau = 1$ the temperatures of the facula and the photosphere are equal. For $\tau > 1$ the temperature below the facula becomes considerably less than the temperature of the photosphere surrounding the facula.

Recently Mustel' and Mitropol'skaya have established that facular regions (regions of facula clustering) act as sources of slow corpuscular streams moving (at the distance of the earth) with velocities of only a few hundred kilometers per second. These corpuscular streams are responsible for the recurring variations in the earth's magnetic field.

§12. Sunspots. Sunspots were discovered by Galileo, using the newly invented telescope, and then independently by Fabricius in 1610. According to various chronicles, large spots visible to the naked eye had often been observed on previous occasions, but Galileo was the first to maintain that they belonged to the surface of the sun.

The formation of a sunspot may be described as follows. First, a dark pore (a small spot 1500 to 3000 km in diameter) appears somewhere in the photosphere. The appearance of a pore is often preceded by the formation, at the same place, of a group of bright facular granules, which then appear to separate, revealing the underlying pore. Once the pore has appeared, it will develop, sometimes within a few hours, into a sunspot with an umbra and a penumbra. Most often a whole group of spots will form, made up of two main spots located at the group edges and a few small spots between these. The westernmost spot in the group, which due to the solar rotation is the first to emerge from beyond the east limb of the disk, is called the leading spot. The small spots within the group soon disappear. Then, toward the end of their lifetime, the large spots break up into several small spots which gradually shrink and finally are lost against the granulation field. Throughout the entire lifetime of a spot, its boundary remains sharp and its contrast with the photosphere remains essentially unchanged.

The lifetime of a sunspot varies from several hours to several months; in 1840–1841, a sunspot was observed which existed for over 18 months. The largest spot was observed in 1858, and had a diameter of 230,000 km. Spots measuring more than 40,000 km are visible to the naked eye. Large spots generally represent a group of several spot centers surrounded by a common penumbra and separated by photospheric "bridges." Figure 14

shows a large spot photographed on 16 July 1905 at Pulkovo by Ganskii. Some spots develop very rapidly. For example, on 24 July 1865 Secchi observed a spot 56,000 km in diameter which formed from three pores which had appeared on the previous day. Then, on the next day, the spot became highly elongated in one direction, reaching a length of 108,000 km.

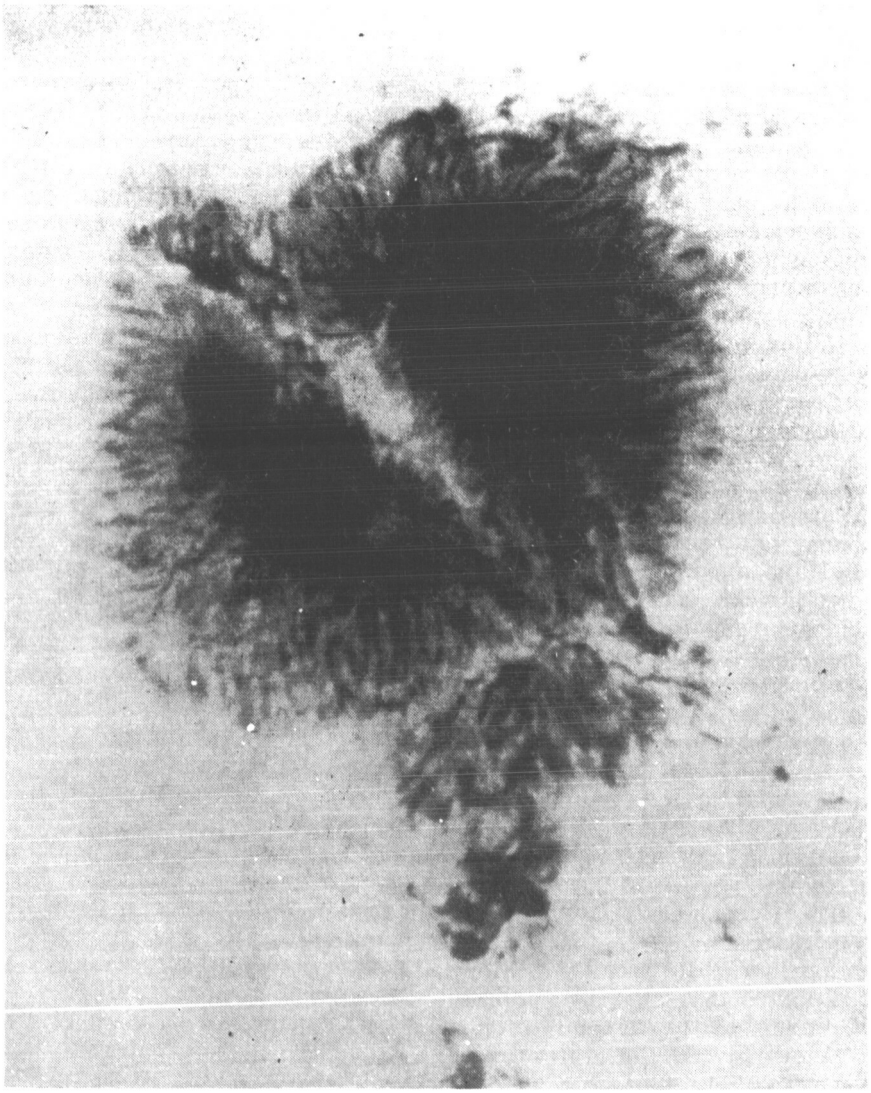


FIGURE 14

The mutual distances between the spots in a group do not vary much. According to Greenwich solar observations over many years, and also according to the observations of Ganskii, the rate of change of these relative distances is some tenths of a kilometer per second. Recently, Gnevysheva

discovered a tendency toward rotational motion in some spot groups. Finally, Brunner has shown that the inclination of the spot-group axis (the line joining the two main spots) to the solar equator decreases systematically with the development of the group.

The granules in the penumbra usually have their major axes oriented parallel to the spot radius, which gives the penumbra an appearance somewhat like a thatched roof. According to some astronomers the inner parts of the penumbra are slightly brighter than the exterior, but other observers do not confirm this phenomenon. The phenomenon may quite possibly be transient in nature and thus may be attributed to the same cause as the appearance of photospheric "bridges" over the spot umbra, these bridges being almost always brighter not only than the penumbra but also than the surrounding photosphere.

According to the Greenwich photographic observations for 1889 to 1923, the area of the umbra constitutes on the average only 0.165 of the total spot area. The spot center appears dark only by contrast. However, it is very difficult to determine the brightness of the umbra relative to that of the photosphere (which is taken as unity), since the light of the sky (the solar aureole) and the light in the instrument lens are the most scattered at the solar disk. Image fluctuations due to atmospheric agitation also have a pronounced effect on measurements. Consequently, all data on the brightness of the umbrae of small sunspots are unreliable and only the results obtained for large spots deserve attention.

The latest data on the ratio (B_s) between the brightness of a spot umbra and the photospheric brightness are those obtained by T. V. Krat from accurate spectroheliographic and spectrographic observations in spectral regions which are virtually free of absorption lines (Table 1).

TABLE 1

λ, Å	4895	5358	6687	8250
B_s	0.144	0.197	0.278	0.350

Sitnik, who carried out sunspot observations using a thermal integrator (see Vol. I, §164), has shown that B_s is independent of the position of the spot on the solar disk. In other words, the contrast between the spot and the surrounding photosphere remains constant regardless of whether the spot is near the center or near the limb. Sitnik's observations refer to the integrated radiation from the spots and the photosphere.

In 1774 Wilson, in Glasgow, observed the following phenomenon. For a circular spot located near the center of the solar disk, the umbra and penumbra are concentric. However, the closer such a spot is to the limb, the greater will be the displacement of the umbra center from the penumbra center; the umbra center is shifted toward the center of the solar disk regardless of whether the spot is situated to the west or to the east of the disk center. Wilson explained this by assuming that the level of the umbra is lower than that of the penumbra, namely that a spot is a depression in the photosphere. He estimated the (average) spot depth as 2000 km. Later, Chevalier used an extensive series of observations at Zô-Sè Observatory

to give a somewhat different estimate of the spot depth, namely 750 km (1"). Other observations have on the whole confirmed Wilson's assumption, but only statistically. The problem is that this phenomenon can be observed only for perfectly circular spots, whose penumbral rims are of equal thickness on all sides. Real spots, however, can never satisfy this condition exactly. It might appear that insofar as the thickness of the photosphere is less than 400 km Wilson's phenomenon is impossible, since spots, faculae, and granulation should exist on the same level. Such a conclusion, however, is wrong. The thickness of the photosphere, as determined from the degree of its opacity, has no direct relation to the profiles of the surfaces of equal absorptivity. Funnel-shaped spots may originate because the photospheric gas in the given region has a lower absorptivity, so that the "lower boundary" of the photosphere appears to recede further down.

In order to facilitate the description of sunspot groups, a 9-class system was worked out at the Zurich Observatory. The classes, based on the external appearances of the groups, represent different stages of group development. We will now give a brief description of the Zurich classes.

A. A group consisting of a single small spot or a small group of small spots. Such a group occupies an area of 2 or 3 square degrees of solar surface. Groups in this category show no structural regularities; the spots have no penumbra and are usually short-lived.

B. A binary (double) group of two small spots without penumbrae. The group axis is oriented approximately along a solar parallel. Still smaller spots and pores have a tendency to concentrate at the western and eastern edges of the group.

C. A binary group, like B, but in which at least one of the spots has a penumbra.

D. A binary group in which the two primary spots have penumbrae.

E. A large binary group of complex structure. Each of the spots has a penumbra. There are many small spots between the main spots. The group has a longitudinal extension of at least 10°.

F. A very large binary group or composite group (one including several large spots). It has a longitudinal extension of at least 15°.

G. A large binary group without small spots or pores. Its longitudinal extension is at least 10°.

H. A single spot with penumbra, sometimes of complex structure. The spot diameter $> 2°.5$. Small spots and pores may sometimes be distributed around it.

I. Similar to H. The spot is usually circular with a diameter $< 2°.5$.

The Zurich classification of spot groups is very approximate and is purely descriptive. However, it is widely used in practice in the Solar Survey, since it covers almost all the possible configurations of spot groups.

§13. Solar rotation, as determined from spots and faculae. After it appears at the eastern limb of the sun, a spot moves across the disk from east to west due to the solar rotation. Seven days later the spot crosses the central meridian, and fourteen days after its appearance it disappears behind the western limb. Many spots (and spot groups) appear several times in succession, and they are called recurrents. The rotation period of a sunspot relative to the central meridian is about 27 days. Because of the inclination of the axis of solar rotation to the plane of the ecliptic, a

sunspot describes an approximate elliptic arc over the solar disk. The eccentricity and the position angle of the axis of this ellipse change during the course of the year, due to the revolution of the earth about the sun and due to the inclination of the solar equator to the ecliptic. On about 6 June and about 7 December, the spots seem to move in straight lines; between 6 June and 7 December the spot trajectory is convex toward the south, and during the rest of the year it is convex toward the north. The position angle of the northern end of the solar-rotation axis, the heliographic latitude of the disk center, and also the heliographic latitude in the Carrington system (see Vol. I, Part V) are listed in astronomical almanacs.

The angular velocity of sunspots (the rotation velocity of the sun) is different at different latitudes, as first demonstrated by Carrington. The angular velocity ξ is generally specified as the angle through which the sun rotates relative to the stars during one mean solar day. Faye was the first to suggest the following formula for ξ as a function of φ (the heliographic latitude):

$$\xi = a - b \sin^2 \varphi, \tag{8}$$

where a is the angular velocity at the equator and b is a coefficient called the equatorial acceleration (it would be more accurate, of course, to call it a deceleration of rotation at higher φ). On the basis of Carrington's observations for 1853–1861, Faye found that

$$\xi = 14^\circ.37 - 3^\circ.10 \sin^2 \varphi. \tag{9}$$

A very thorough study of the solar rotation using sunspots was carried out by Maunder, who based his work on the Greenwich photographs of the sun for 1879–1901. Maunder found that

$$\xi = 14^\circ.44 - 2^\circ.01 \sin^2 \varphi.$$

The most accurate values of constants a and b, however, are those obtained at Greenwich from studies of long-lived spots (lifetimes of 25 days or more), using photographs for the period from 1879 to 1933. Such spots are usually circular and appear as the leading spots in groups. A total of 579 spots were investigated, at latitudes up to ±30° and over four solar-activity cycles. The sidereal-rotation period of the equatorial zone of the sun is 25 days, and it increases to 26 days at 30° latitude. Recently, Newton and Nunn /5/ extended the Greenwich computations up to 1944, and the final formula is

$$\xi = 14^\circ.38 - 2^\circ.77 \sin^2 \varphi. \tag{10}$$

Wilsing at Potsdam in 1888 was the first to study the solar rotation on the basis of faculae. However, he arrived at an erroneous conclusion, that the sun rotates as a rigid body. The inaccuracy of Wilsing's conclusion was demonstrated in 1891 by Belopol'skii, using Pulkovo-Observatory photographs. The first reliable determinations of a and b for faculae were those of Stratonov, who used measurements of the Pulkovo photographs for 1891–1894 to obtain the formula

$$\xi = 14^\circ.56 - 2^\circ.78 \sin^2 \varphi.$$

At Greenwich solar photographs for the period from 1888 to 1923 were used for this same purpose. In this case the procedure for determining a and b was somewhat different. First, photographs were chosen which were representative of the situation near the solar-activity minimum, and then small isolated facular regions were identified on these photographs for the duration of one complete solar rotation. The results of these investigations were in good agreement with the data of Stratonov, and the formula obtained was

$$\xi = 14^\circ.54 - 2^\circ.81 \sin^2 \varphi. \qquad (11)$$

A comparison of formulas (10) and (11) shows that the equatorial acceleration is the same for sunspots and for faculae. On the whole, however, it appears that the velocity ξ calculated from faculae is higher than the value calculated from sunspots. Let us note, though, that when determining the velocity of solar rotation from sunspots and faculae we are actually dealing not with a geometrical but with a statistical problem. Even at a single latitude spots have various velocities of rotation about the solar axis due to their proper motions, which show up as changes in the mutual distances of spots in a group and changes in the motions of spot groups in the photosphere. It has been established, however, that the latitudinal (proper) motions of sunspots average out to about zero (that is, they either rotate or move at random).

Although spots are relatively stable, faculae, on the other hand, which are made up of strings of facular granules, are quite emphemeral and cannot be regarded as formations which are traceable for considerable lengths of time. Faculae remain in more or less the same place for a fairly long time, however, so that facular granules which have decayed can reform in the same place as before, with the result that the process of facula formation is comparatively localized in the photosphere. That the solar rotation calculated from faculae is higher than that calculated from spots should be considered as due to a shifting of the facula-formation foci in the direction of solar rotation, relative to the spot groups.

The exact position of the solar-rotation axis is determined by analyzing the apparent spot trajectories over the solar disk; long-lived recurrent spots are particularly useful in this respect. The first determination of the inclination angle i of the solar equator to the plane of the ecliptic and the first determination of the ecliptic longitude Ω of the ascending node of the solar equator were made by Carrington. He obtained

$$i = 6^\circ 58' \text{ and } \Omega = 74^\circ 22' \quad (1850).$$

In 1913 Dyson determined these values from photographic observations made at Greenwich from 1874 to 1912. He obtained

$$i = 7^\circ 10'.5 \text{ and } \Omega = 73^\circ 46'.8 \quad (1850).$$

In 1915 Orlov and Tsimmerman determined i and Ω from photographs of observations at Odessa from 1894 to 1911, to obtain

$$i = 7^\circ 6' \text{ and } \Omega = 73^\circ 48' \quad (1850).$$

Finally, Servajean recently obtained the following values:

$$i = 7°0' \text{ and } \Omega = 74°2' \quad (1930).$$

The latter determinations show that Carrington's values are quite reliable and can be used in most studies.

§14. **Special features of the line spectra of spots and faculae.** The spectra of spots and faculae are usually studied by means of a comparison with the spectrum of the adjacent photospheric regions. Such comparisons are made easy by the fact that spots are, as a rule, spectrographed by centering the spectrograph slit onto a spot on the solar image in such a way that the spectrogram simultaneously records the spectra of the spot umbra, the penumbra, and the adjacent regions of the photosphere. Faculae are spectrographed in the same way. A special photographic atlas of the spectra of sunspots and the photosphere, covering the region from 4600 to 7200 Å, has been compiled at Mount Wilson Observatory.

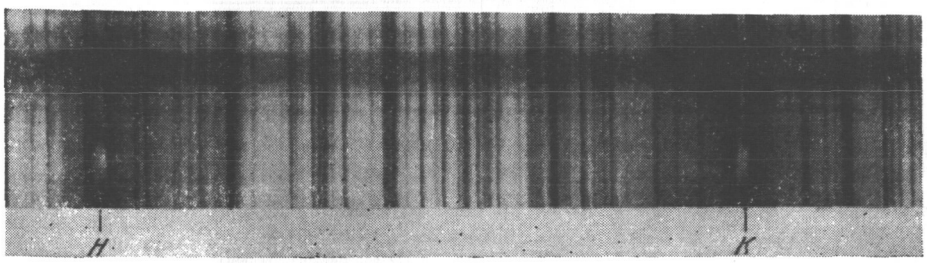


FIGURE 15

In comparison with the spectrum of the photosphere, the lines of ionized atoms of various elements are attenuated in the spot spectra. Figure 15 shows part of the spectrum of a sunspot and the surrounding photosphere (photographed at Pulkovo Observatory with a horizontal solar telescope and a diffraction spectrograph). Several molecular bands are present in the spot spectrum but are not visible in the spectrum of the photosphere. The spectra of different spots have the same appearance, which indicates that they have identical physical structures.

Let us define active atoms as those which absorb light and cause a variation in the emitted flux at a given frequency (the effect of this [selective] absorption at great depths is weakened by the continuous absorption in the overlying strata). Then, we must assume that the line intensities for a given element in the spectrum depend on the number of active atoms occurring in various energy states in a photospheric column 1 cm^2 in section (henceforth referred to simply as the number of atoms), parallel to the observer's line of sight. This number of atoms depends in turn on the temperature, on the gas pressure, and on several constants characterizing the atomic energy state (excitation, ionization). In 1947, T. V. Krat /6/ showed that the line intensities for spot spectra can be computed on the basis of the hypothesis of so-called local thermodynamic equilibrium, using the formulas of Boltzmann and Saha. Therefore, a comparison of the equivalent widths of the same spectral lines in the photosphere and spot spectra will make it possible to determine the temperature of the spots and

the ratio of the electron pressures (and consequently the total pressures as well) in the spots and in the photosphere. However, it should be kept in mind that the growth curves, which express the equivalent line widths as functions of the number of active atoms, are different for spots and for the undisturbed photosphere. Line comparison is thus justified only within a single growth curve; for example, this may be a comparison of the lines of neutral and ionized elements or a comparison of the lines of a single element but which originate at different energy levels (with different ground-level excitation potentials). It should also be kept in mind that the lines in sunspot spectra show Zeeman splitting, since the magnetic fields of sunspots are considerable. In order to avoid having to take the Zeeman effect into consideration, it is possible to use only the very wide lines (the hydrogen lines, the lines of ionized and neutral calcium, the green magnesium triplet, the D_1 and D_2 sodium lines, etc.), whose equivalent widths are practically unaltered by the Zeeman effect.

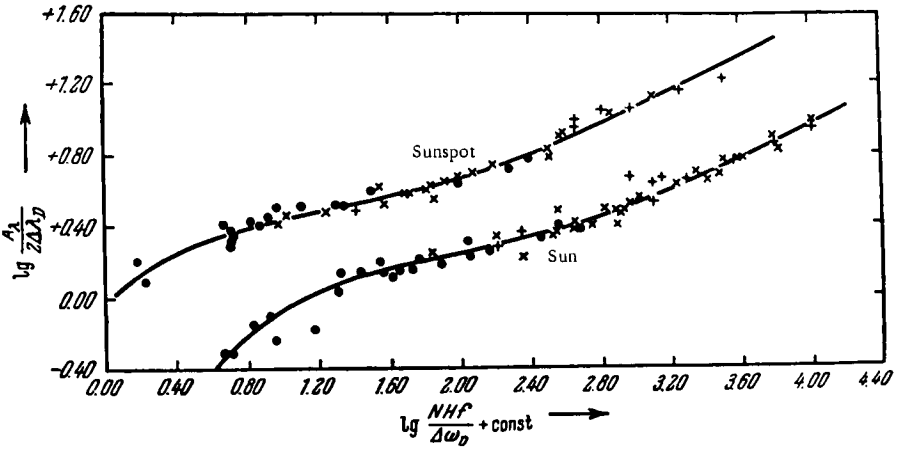


FIGURE 16

The first study of this kind, based on naked-eye estimates of the line intensities in sunspot and photosphere spectra, was carried out by Moore in 1931. More reliable data were obtained by ten Bruggencate and von Klüber /7/ and by T. V. Krat /6/, using exact methods of photographic spectrophotometry and taking into account the various systematic errors due to the earth's atmosphere and to instrument effects. Ten Bruggencate and von Klüber plotted separate growth curves for the photosphere and the sunspots (Figure 16), in order to determine the ratios between the numbers of atoms in given energy states. For these curves, they considered only the large spots, whose spectral lines are not distorted by atmospheric vibration of the solar image. Since on photographs of the solar disk all spot umbrae (with the possible exception of pores) show the same contrast with the photosphere, it is clear that such spectrophotometric studies of the large spots will ensure the most reliable results.

The growth curves plotted by ten Bruggencate and von Klüber for the normal photosphere and for sunspots are shown in Figure 16. The abscissa

is the logarithm of a quantity proportional to the number of oscillators (the product of the number of atoms times the oscillator strength), while the ordinate is the logarithm of the equivalent width (A_λ) divided by $2\Delta\lambda_D$, where $\Delta\lambda_D$ is the so-called Doppler line width.

Recently the studies of ten Bruggencate and von Klüber were repeated by Abhyankar and Ramanathan /8/ using Cr, Ti, and Fe lines and by Kornilov using Ti, Fe, V, Ti^+, and Fe^+ lines. Their growth curves came out very close to those plotted by the earlier authors.

Since the growth curve illustrates graphically the equivalent line width as a function of the number of atoms, therefore if these curves are specified for the spot umbra and for the photosphere then we can obtain the desired ratio of the numbers of atoms of some element (the curves are plotted for neutral iron and for titanium) in the spot umbra to the corresponding numbers of atoms in the photosphere.

Let us first write Boltzmann's formula separately for the spot (s) and the photosphere (p):

$$\lg\left(\frac{n_{i,p}}{n_{0,p}}\right)=C-\chi_i\theta_p, \tag{12}$$

$$\lg\left(\frac{n_{i,s}}{n_{0,s}}\right)=C-\chi_i\theta_s, \tag{13}$$

$$\theta_p=\frac{5040^\circ}{T_p}, \quad \theta_s=\frac{5040^\circ}{T_s}.$$

Here the term C contains the statistical weights of the states, n_0 is the number of atoms per cm^3 located in the ground energy state, n_i is the number of atoms in the ith energy state, and χ_i is the excitation potential. Subtracting (12) from (13), we have

$$\lg\frac{n_{i,s}}{n_{i,p}}=\lg\frac{n_{0,s}}{n_{0,p}}-\chi_i(\theta_s-\theta_p). \tag{14}$$

If we denote the difference $\theta_s-\theta_p$ as $\Delta\theta$ and write relations such as (14) for the different energy states of a given atom, then we obtain a system of equations in two unknowns, namely $\Delta\theta$ and $\lg\frac{n_{i,s}}{n_{i,p}}$. If we take into account that for weak lines the n_i values are proportional to the equivalent line widths, then it is easy to find $\lg\frac{n_{0,s}}{n_{0,p}}$ and from this the unknown $\lg\frac{n_{0,s}}{n_{0,p}}$. Once the numbers of neutral atoms and ions in the ground state are known, we can apply Saha's formula to calculate the ionization of the gas. Then, for the spots and for the normal photosphere we will obtain two equations:

$$\lg\left(\frac{n_0^+}{n_0}\right)_s+\lg(p_e)_s=-\chi_0\theta_s+\frac{5}{2}\lg T_s+C', \tag{15}$$

$$\lg\left(\frac{n_0^+}{n_0}\right)_p+\lg(p_e)_p=-\chi_0\theta_p+\frac{5}{2}\lg T_p+C', \tag{16}$$

where C' is a constant depending on the statistical weights of the ionized and neutral states of the atom, and $p_{e,s}$ and $p_{e,p}$ denote the electron pressures in the spot and in the photosphere. By subtracting (16) from (15), we obtain

$$\lg\frac{n_{0,s}^+}{n_{0,p}^+}+\lg\frac{p_{e,s}}{p_{e,p}}=\lg\frac{n_{0,s}}{n_{0,p}}-\chi_0\Delta\theta+\frac{5}{2}\lg\frac{T_s}{T_p}. \tag{17}$$

When $\Delta\theta$, $\frac{T_s}{T_p}$, $\frac{n_{0,s}^{+}}{n_{0,p}^{+}}$ and $\frac{n_{0,s}}{n_{0,p}}$ are known, it is easy to determine the ratio $\frac{p_{e,s}}{p_{e,p}}$. According to ten Bruggencate and von Klüber, $\Delta\theta = 0.33$, $T_p = 5040°$, and $T_s = 3800°$. Thus, by applying equation (17) for Ca II (H and K lines) and Ca I (line 4227 Å), T. V. Krat obtained

$$\lg \frac{p_{e,s}}{p_{e,p}} = -1.60 \text{ and } \frac{p_{e,s}}{p_{e,p}} = \frac{1}{40}.$$

Mustel' maintains that these numbers characterize $p_{e,s}$ for large spots only. For small spots the ratio $\frac{p_{e,s}}{p_{e,p}}$ may be closer to unity. Unfortunately, it is not yet possible to perform exact spectral-line photometry for small spots, because their umbrae are blurred by vibrations of the image on the spectrograph slit. At any rate, it is clear that $p_{e,s} \ll p_{e,p}$. This is mainly due to the fact that over a spot umbra the gas is more rarefied than over the hotter regions of the photosphere. The lower ionization of the gas in the sunspot also has some effect.

Analogous studies of the Fraunhofer spectra of faculae were carried out by T. V. Krat /10/ and later by Mitropol'skaya /11/. As we have seen (see §3), a facular granule (as well as the facula proper, which is an accumulation of several such granules) can best be regarded as a thin layer of gas on the photosphere surface which is over 1000° hotter than the surrounding photospheric gas. Measurements of the contrast between the faculae and the photosphere at various wavelengths support this conclusion; the contrast increases for shorter wavelengths. Quantitatively this contrast is usually small, but the facula spectrum is blurred by the photospheric spectrum because of atmospheric vibration of the solar image. The image-vibration amplitude for low vibration frequencies is almost always much greater than that for high frequencies. Thus, short-exposure photographs come out sharper than long-exposure photographs, since the latter are blurred by low-frequency vibrations.

Recently Mel'nikov and Zhuravlev /12/, who assumed a facula to be transparent in the continuous spectrum with its brightness superimposed onto that of the undisturbed photosphere, obtained spectrophotometric facula temperatures 1000 to 3000° in excess of the temperature at the center of the solar disk.

T. V. Krat obtained the equivalent line widths for Ca 4227 Å and Ca II K; Fe I 5049.83; 5051.64; 5109.66; 5110.41; 5133.70 Å and 5139.26, 5139.48 Å; Fe II 4923.93 and 5018.45 Å; Ti I 5007.22, 5022.88, 5024.85, 5025.57 Å; and Ti II 5154.08 Å. It was noted that the lines of ionized Fe and Ti in the spectra of faculae are stronger than in the photospheric spectrum, while, on the other hand, the lines of neutral Fe and Ti are weaker. This fact definitely indicates a higher ionization of metals in the facular granules. The calcium lines are somewhat anomalous in their behavior. Both the K line and the Ca 4227 line are weaker in the facula spectrum, the degree of attenuation being greater for the 4227 Å line. According to T. V. Krat, this may be a consequence of the fact that a large proportion of the calcium atoms are ionized to the Ca III state. Mitropol'skaya, who agrees with this conclusion, goes on to observe that the increased calcium ionization may be due to ultraviolet radiation beyond the limit of the Lyman series. However, this has not, in our opinion, been proved definitely. The problem is

that the ionization potential of Ca I is $\chi_0 = 6.09$ V, while that of Ca II is $\chi_1 = 11.82$ V. Under such conditions the ionization may be increased by electronic impact, a factor which becomes significant for a rise in T of 1000 , and to a certain extent it can even be increased by photospheric radiation below the Lyman limit.

A growth curve for faculae was first constructed by Mitropol'skaya. She also studied in detail the profiles of the H_α and H_β lines of hydrogen in the spectrum of the faculae. Simultaneously with Ismailov, she concluded that the line profiles in this case are determined by two gas layers of different temperature, a low, cooler layer and an upper, hotter layer. This conclusion, moreover, is consistent with the present-day concept of facular granules (§11). However, this factor complicates the derivation of the quantitative characteristics of faculae considerably: the electron pressure, the excitation and ionization temperatures, etc. The problem is very difficult from a theoretical point of view as well, since the conditions for local thermodynamic equilibrium are not observed in a thin heated gas layer, with the result that the formulas of Boltzmann and Saha do not apply.

§15. Magnetic fields of spots. The general magnetic field of the sun. The first reliable information on solar magnetic fields was obtained from the Zeeman splitting of lines in the spot spectra. If the line of sight is perpendicular to the magnetic lines of force surrounding a light source, then the spectral lines are split into three components: a central component which does not shift in wavelength and two components located symmetrically one on each side of the line center. The light of all these line components is linearly polarized; for the central component it is polarized parallel to the lines of force and for the lateral components it is polarized perpendicular to them. If, on the other hand, the line of sight is parallel to the magnetic lines of force, the central component is not observed and the spectral line appears to be made up just of two components. The light of these components is circularly polarized in opposite directions. The distance between the outer components in either case is proportional to the magnetic-field intensity, as the following formula shows:

$$\Delta\lambda = 4.67 \cdot 10^{-5} H\lambda^2, \qquad (18)$$

where λ and $\Delta\lambda$ are in centimeters and H is in oersteds. In reality, the splitting is more complicated. The field intensity H can be measured from the lines of simplest form, which to a first approximation coincide with the classical Zeeman triplet and which also show the greatest splitting. The amount of splitting is determined by the so-called Landé g-factor. When g is introduced into (18), we obtain

$$\Delta\lambda = 4.67 \cdot 10^{-5} Hg\lambda^2.$$

For some lines which are particularly convenient for determining H, Table 2 (compiled by von Klüber) gives the values of $10^{10}\, g\lambda^2$.

From 1917 on, the magnetic-field strengths and polarities for all the visible spots have been determined daily at Mount Wilson Observatory. The polarity is determined from the intensity ratio for the split line components. If the relative intensities of the three components are i_v (the short-wave "violet" component), i_c (the central unshifted component), and i_r (the long-wave "red" component), then for all i we obtain the following functions

of the angle (γ) between the line of sight and the intensity vector of the magnetic field:

$$i_v = \frac{1}{4}(1-\cos\gamma)^2, \quad i_c = \frac{1}{2}\sin^2\gamma, \quad i_r = \frac{1}{4}(1+\cos\gamma)^2,$$
$$i_v + i_c + i_r = 1.$$

From the attenuation of the oppositely polarized components i_v and i_r, we can determine the sign of the field intensity (whether it is "north" or "south" polarity). The magnetic fields of sunspots were previously studied at Mount Wilson using a diffraction spectrograph with a Nicol prism and a quarter-wave plate ahead of the slit. Recently, however, in order to obtain a higher resolving power and also a higher accuracy of the results, von Klüber used equipment incorporating a Fabry-Perot interferometer, in which the diffraction spectrograph just plays the role of the primary monochromator. The average error in H in this case is about ±25 oersteds.

TABLE 2

Line	Element	Rowland intensity		$10^{10} g\lambda^2$
		photosphere	sunspot	
5131.478	Fe	2	3	66
5247.576	C	3	5	69
5250.218	Fe	2	3	83
6173.348	Fe	5	5	95
6258.578	V	1	2	132
6302.508	Fe	5	4	100
6733.162	Fe	1	0	114
8468.417	Fe	2	4	180
10288.87	Si	—	—	212

The most important result of the Mount Wilson observations was the derivation of the law of the magnetic polarity of sunspots, discovered by Hale. It was established that spot groups can be divided into the following three classes (Halc and Nicholson /14/).

a) A unipolar group includes either a single spot or a group of spots which all have the same magnetic polarity.

b) A bipolar group includes, in the simplest case, two spots (a binary group) with opposite magnetic polarity. The axis of the group (the line joining the spots) makes only a small angle with the solar parallel. Often instead of two spots we observe two groups of small sunspots, which constitute the leading and following components of the group and which, like the individual spots, have opposite magnetic polarity.

c) A complex spot group consists of spots of different polarity, arranged very irregularly.

Strictly speaking, unipolar groups either are still undeveloped or else are very old bipolar groups in which one of the spots has been replaced by a region of opposite magnetic polarity. Hale called such regions "invisible sunspots" and detected their presence by the presence of the magnetic field. In some cases the magnetic fields presumably occupying the places of the second spots in bipolar groups had intensities of about 200 oersteds (and not higher than 500 oersteds). Complex groups are in all probability heterogeneous formations which originate due to the overlapping of several bipolar

groups. The bipolar group is the basic and most characteristic formation among sunspot groups.

The study of the magnetic fields of bipolar groups led Hale to discover the law of magnetic-polarity reversal in sunspots. It was established that within one (11-year) solar-activity cycle all the leading spots (that is, those going ahead of the group during solar rotation) have the same polarity. In other words, all the bipolar groups have the same longitudinal orientation. When a new cycle commences, however, this orientation becomes reversed, so that if in the preceding cycle all the leading sunspots were of north polarity then in the new cycle they will all be of south polarity. About 98% of all the bipolar groups obey this law. Therefore, the initial magnetic-field distribution in the sunspot groups on the solar surface is restored only after 22 or 23 years.

The presence of strong magnetic fields in sunspots also affects their curves of growth. As shown by Mel'nikov and Zhuravlev /13/, magnetic splitting intensifies the absorption lines; thus, if this strengthening is interpreted as an effect of Doppler "turbulent" motions, then the "turbulence" in the gas is also increased. It is clear, however, that the "turbulence" is in this case purely fictitious.

Also of great interest is the general magnetic field of the sun. The first data on this field were obtained by Hale in 1912. His observations led him to conclude that the sun may be regarded as a magnetized sphere (dipole). According to Hale's data, the magnetic poles of the sun are close to the poles of the rotation axis, the north magnetic pole being located in the northern hemisphere, just as on the earth.

The studies made at Mount Wilson Observatory in 1914 and 1916 by Hale, Seares, van Maanen, and Ellerman gave the following results:

the inclination of the magnetic axis to the solar rotation axis is $i = 6^{\circ}.0 \pm 0^{\circ}.4$;
the rotation time for the magnetic pole is $T_m = 31.52 \pm 28$ days;
the field intensity at the poles is $H_{pol} = 50$ oersteds.

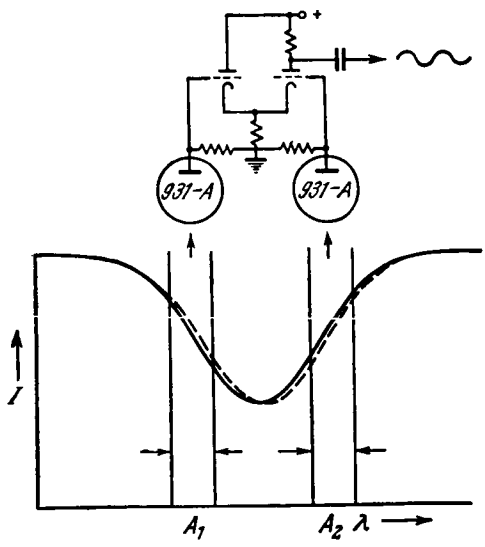


FIGURE 17

Later Hale reduced the value of H_{pol} to 20 oersteds /14/. More recent studies, however, have not confirmed the results of the Mount Wilson astronomers. Thiessen, using a Fabry-Pérot interferometer, found H_{pol} to be 1.5 oersteds in 1949 and 2.5 oersteds in 1951. Von Klüber obtained the value $H_{pol} = 0$, with a possible upper limit of 1 oersted. Even more recently, careful investigations of the weak magnetic fields on the sun have been carried out by Babcock and Babcock /15/ at the Hale Solar Laboratory in Pasadena (California). These investigators employed a vacuum spectrograph with a diffraction grating 13 × 20 cm in ruled area and with 600 grooves per mm, using fifth-order green light. The dispersion was 11 mm/Å and the resolving power was 600,000. The splitting of the 5250 Å line was recorded as follows using a modulation photoelectric photometer [detector]. The double slit was centered onto the wings of the line profile (Figure 17). Photomultipliers were placed opposite each slit (A_1 and A_2) in order to record not the total luminous flux but only the differences in the fluxes passing through the two slits. For a line with Zeeman splitting this difference (provided the observations are made in polarized light) will naturally be a function of H. A Nicol prism and an electro-optical retardation plate made of ammonium dihydrogen phosphate were used as an analyzer. The solar image, which was 40 cm in diameter, was allowed to drift across the spectrograph slit. The effect of the magnetic field, which increases the splitting of the line, showed up as a shift in a point of light on an oscilloscope screen which recorded the photocurrent. The oscilloscope screen was photographed on a moving film, which showed the profiles of the magnetic-field sections in the direction of motion of the solar image across the slit.

Some recordings of H are shown in Figure 18, the vertical distance between lines being one oersted. The accuracy of the given oscilloscope recording is ±0.1 oersted (the accuracy depends on the slit width, that is, on the signal amplitude). A superficial examination of the Babcock chart indicates that no general magnetic field of the sun exists. However, in both the northern and southern solar hemispheres, weak local fields appear near the poles, which have intensities up to one oersted and like polarities. Near the equator the weak magnetic fields are generally opposite in sign. These facts enable us to interpret the general magnetic field of the sun as a statistical mean of many weak local fields, which for some still-unknown reason are oriented by the solar rotation. For this mean field $H_{pol} = 1$ oersted. There is some reason to believe that H for the mean field varies during the 11-year solar cycle. Of particular interest, in view of their relation to geomagnetic phenomena, are the weak but extensive (10 to 30° in longitude) unipolar fields, which possess high stability. The maximum value of H for these fields does not exceed 1 oersted.

The Babcock charts show quite well the magnetic fields averaged over areas occupied by strings of faculae. These fields often exceed 10 oersteds (and may reach 20 or 30 oersteds).

Recently, magnetograph observations have been carried out at the Crimean Astrophysical Observatory /16/ and at Pulkovo Observatory /17/, using equipment similar in principle to the magnetograph setup of Babcock and Babcock. On the whole, these have confirmed the conclusion concerning the irregular structure of the magnetic field on the solar surface; however, since the recording was made using a narrow slit (up to several arc seconds in width),

it was found that very local magnetic fields on the sun have H values which are several times greater than those obtained by Babcock and Babcock using their wider slit. Obviously, the magnetic field possesses a comparatively fine structure which is blurred, or more precisely is averaged out, in wide-slit recordings. Also very interesting are the results of Stepanov, who recorded the magnetic fields of faculae and constructed charts of the H isolines for facular regions. Stepanov and Petrova found that, although the local maxima of H in facular regions are higher than those in undisturbed regions of the sun (they may reach several tens of oersteds), still the positions of these maxima do not coincide with the centers of H_α-line emission for faculae, which form over photospheric faculae. The faculae appear to be situated on the downward part of the H curve, near the maximum, as if the facular emission were actually due to currents induced around a crowding of the magnetic lines of force. For $H < 50$ oersteds, the faculae coincided with the maximum of H. The overall pattern, of course, does not correspond to any simple model. The magnetic field possesses many irregularities (near the maxima), and the distribution of facular granules is even more irregular. The very fact that faculae are divided up into granules denies the existence of regular annular currents in the photosphere and of any other very extended currents as well.

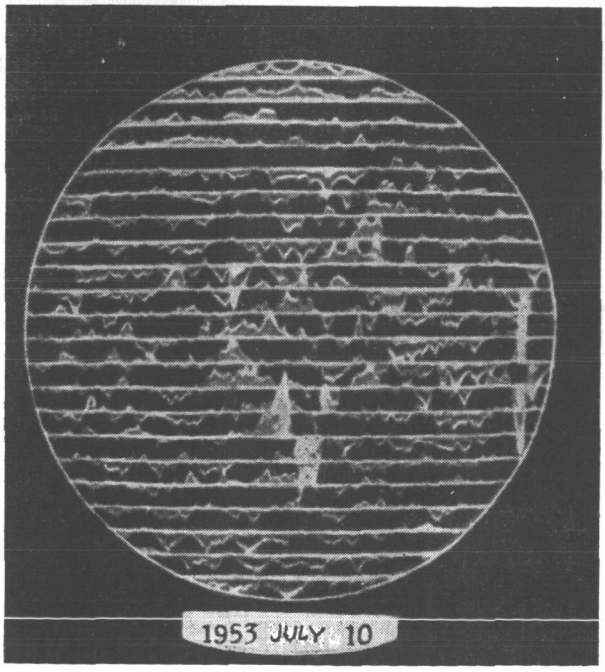


FIGURE 18

§16. Gas motion in sunspots. The Evershed effect. Evershed (1909, 1916) was the first to begin a systematic study of the motion of gas in sunspots, by means of measurements of the radial velocities as reflected in

the spot spectra. The Doppler shift of the spectral lines which he discovered in sunspots is known as the Evershed effect. If a spectrograph slit is oriented along the radius of the solar disk, then the edge of the spot penumbra closer to the disk center shows a Doppler shift toward lower wavelengths, while the opposite edge (the one toward the limb) shows a shift toward higher wavelengths. This shift becomes more pronounced toward the edge of the penumbra, and abruptly disappears immediately upon transition from the penumbra to the surrounding photosphere. The magnitude of the shift increases as the sunspot approaches the limb. The Evershed effect is generally interpreted as an efflux of gas from the spot umbra, which takes place in an almost exactly horizontal direction. The efflux velocity at the outer edge of the penumbra usually reaches 2 km/sec.

St. John (1931) studied the Evershed effect in more detail and established that the phenomenon is actually more complicated than just a simple efflux of gas from the umbra. Such an efflux can be distinguished only according to the weak lines of metals (preferably the Fe lines). For stronger metal lines (Fe, Mg), the Evershed shift is practically zero. This also applies to the Ca 4227 line, which shows a very slight shift, but in a direction opposite to that of the H_δ line; the D_1 and D_2 (Na), b_1, and b_2 (Mg) Fraunhofer lines, however, show a substantial reverse shift. The reverse shift, which corresponds to an influx of gas into the sunspot, is the strongest for the H_γ, H_β, and H_α lines, and particularly in the central cores of the H and K lines (Ca II). The influx velocities for the H and K cores can reach 3 km/sec. The cores of strong Fraunhofer lines are formed in the upper layers of the solar atmosphere, since light emitted by the lower strata actually becomes absorbed due to the opacity of the solar atmosphere to the corresponding wavelengths. The phenomenon of gas efflux from spots thus characterizes the motion of gas in the upper layers. The pattern of the gas flow in sunspots is reminiscent of the terrestrial breezes which blow over the cooling land of islands. The cool air must flow out toward the sides, while warm air descends from above. However, the analogy is not perfect, since the gas motion which ceases at the outer edge of the penumbra does not change into a vertical stream but definitely continues along below the surface of the photosphere.

In 1926 at the Arcetri Observatory (Italy), Abetti once again studied the Evershed effect. On one occasion he observed a small sunspot whose umbra underwent violent changes. In this umbra the efflux velocity was 6.6 km/sec, and various photographs showed that the maximum efflux velocity was observed not at the [spot] periphery but at the edge of the umbra; the velocity decreased nonlinearly with the distance from the umbra.

In 1915 Evershed observed in a certain sunspot a phenomenon which later became known as the Evershed effect of the second kind. When this spot was located on the eastern side of the solar disk, the eastward part of the penumbra had a higher velocity than the westward part. When the sunspot crossed the central meridian, this pattern reversed and the westward part of the penumbra acquired a greater velocity. This phenomenon was observed repeatedly at the Arcetri Observatory. A correct explanation was given by Moncinelli. According to Moncinelli, the efflux of the gases is not horizontal, but slopes downward, making an angle of from 2 to 29° with the surface of the sun. Thus, it follows that the spot penumbra represents

during the greater part of the sunspot lifetime the magnetic field remains constant, whereas the spot area, once it has reached a maximum, steadily decreases. Therefore, it looks as if the magnetic field does not form together with the sunspot but rather appears to emerge to the surface and then to sink down again below the photosphere.

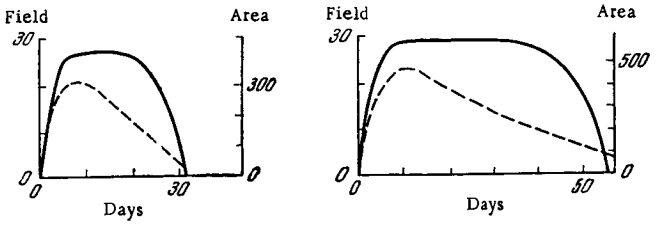


FIGURE 19

At the center of the sunspot the lines of force are almost vertical. At the inner edge of the penumbra they slope outward, making about a 25° angle with the vertical. This angle (θ) varies with the distance (r) from the spot center as follows:

$$\theta = 90° \frac{r}{b}, \tag{19}$$

where b is the radius of the outer boundary of the penumbra. The field intensity H depends on the maximum sunspot area (during its whole lifetime) and may be as high as 4000 oersteds for the largest sunspots. Inside the sunspot, according to Broxon (1942) and Thiessen (1953), H is related to H_m (the field intensity at the center) by the expression

$$H(r) = H_m\left(1 - \frac{r^2}{b^2}\right).$$

Near the solar limb the field intensity H of a sunspot decreases, which may be due to the variation of H toward the upper regions of the atmosphere. In this case the effect is completely analogous to limb darkening. In both cases a raising of the effective level of the emitting layer is caused by an increase in its optical thickness due to the oblique (relative to the horizontal plane) path of the rays. The quantity H actually does decrease rapidly with altitude. For large sunspots Houtgast and van Sluiters obtained $\frac{dH_m}{dh} =$ 5.7 oe/km. On the other hand, the divergence of the lines of force observed by Hale and Nicholson sets this gradient at 0.5 oe/km. Moreover, King compared the H values obtained from the strongest and weakest Fraunhofer lines and referred them to the different effective depths of origin, and he obtained $\frac{dH_m}{dh} = 2.5$ oe/km. All these estimates agree roughly, indicating that the sunspot field pattern in the solar atmosphere is approximately that in a vacuum.

Solar gas is a good conductor. Near the surface of the sun, at $T \approx 10^4$ degrees, the conductivity (σ) is about 10^{-8} electromagnetic units. On the

an elevation above the surface of the photosphere, which contradicts the Wilson effect somewhat, but is nevertheless consistent with certain direct observations.

Evershed, St. John, and Abetti also attempted to determine the rotation velocities of sunspots. In these studies, the spectrograph slit was aligned parallel to the radius of the solar disk (for a sunspot near the solar limb). Under such conditions, of course, only the velocity component tangential to the penumbra was determined. The rotation velocity varied in both magnitude and direction for different sunspots. From the K_3 line St. John obtained a rotation velocity of 1.59 km/sec for five sunspots. Abetti (1932), on the basis of observations of 26 spot groups from 1926 to 1930, showed that the direction of sunspot rotation obtained from the metallic lines is generally opposite to the direction of motion of the hydrogen vortexes (see next chapter) which show up on spectroheliograms. In 1934 Calamai found that the gas in sunspots moves along logarithmic spirals.

§17. The nature of sunspots. Most of the sunspot theories which have been formulated are now of historical interest only. Of these the theory of Stepanov /18/ is perhaps the one which is hydrodynamically the most sound. However, even this theory cannot be accepted completely, since it essentially neglects the role of the magnetic fields of the spots, these being of prime importance in all hydromechanical computations. The greatest problem at present is not how to interpret the gas motion in sunspots but rather how to explain the origin of the magnetic fields which specify the physical conditions in sunspots. For example, Biermann /19/ has shown that the temperature drop in a spot can be considered as an immediate consequence of the presence of the magnetic field. An intense magnetic field located below the photosphere can stop the convection almost completely, since the gas masses, which are good conductors, will move along the lines of force of the field. During convective motion, on the other hand, gas currents will inevitably intersect the magnetic lines of force. Consequently, under a sunspot energy is transferred by radiation only. Since the radiative temperature gradient is greater than the gradient in the convective zone, the temperature in a sunspot will naturally be lower than the normal photospheric temperature.

The origin of the magnetic fields of sunspots has still not found a satisfactory theoretical explanation. Gurevich and Lebedinskii /20/ were the first to attribute the origin of the solar magnetic fields to the motion of a conductor. For instance, in particular, the compression of a conductor with "rigid" magnetic lines of force may increase the number of lines of force per unit volume considerably, that is, it may increase the field strength. Large-scale hurricanes apparently occur in the interior of the sun, at depths of 10^4 or 10^5 km below the photosphere, and their mechanical energy is partly converted into the energy of magnetic fields. The mechanism of this conversion, however, is still unknown, and the theory describing this phenomenon has not yet been developed.

The magnetic fields of sunspots are exceptionally stable. In 1946 Cowling /21/ used data on sunspot magnetic fields obtained at Mount Wilson to make a preliminary attempt at describing the development of a spot field. His curves relating the magnetic field to the spot area are shown in Figure 19. The curves are averaged, and nonsystematic daily fluctuations of the magnetic field have been smoothed. It is evident from these curves that

earth this value is close to the conductivity of salt water ($\sigma \approx 10^{-9}$ electromagnetic units). Thus, the gas conductivity is not high but just intermediate. The behavior of individual prominences above sunspots indicates that the sunspot magnetic field may extend to great heights (the prominences often move along the lines of force of the sunspot field). It is still not clear to what extent the variation of H in a sunspot is related to the Evershed effect. There is apparently no direct relation between the two, since the efflux of gas in a sunspot represents a one-way process, whereas the value of H in a spot is subject to considerable fluctuation. Nevertheless, it is almost certain that the variation in H must play an important part during the epoch of most rapid sunspot growth; this variation actually causes the field, and consequently the sunspot itself, to emerge to the surface of the photosphere.

Gas should flow out of a sunspot mainly along the magnetic lines of force. Therefore, if the gas flow slopes downward, below the photosphere, then this must also be the direction of the magnetic lines of force. The existence of the sharp boundary of the penumbra and sunspot, however, is still puzzling. Why should all apparent motion cease at this boundary, despite the fact that the sunspot magnetic field can extend over greater distances ? No theoretical answer has so far been given for this question.

§18. The cyclic variation of photospheric phenomena. Although the number of sunspots observable on the solar disk varies greatly from day to day, the average yearly number of sunspots behaves differently. If we calculate the yearly average of the number of sunspots observed each day and plot a curve of these sunspot numbers over many years, then we obtain a smooth curve. The German amateur astronomer Schwabe (a druggist by profession) was the first to establish, from his observations in 1843, that the number of sunspots shows a definite periodicity. Schwabe estimated the period between two maxima at 10 years. Wolf of Zurich became interested in Schwabe's discovery and studied all the sunspot observations from the time of Galileo to his own period. Wolf confirmed Schwabe's results and established an average period of 11 years. In order to characterize sunspot activity properly , Wolf introduced the concept of the "relative sunspot number" (Wolf number). If g denotes the number of sunspot groups observed on any given day and f is the number of individual spots on that day, then the relative spot number is defined by the empirical formula

$$R = k(10g + f), \tag{20}$$

where k is a coefficient depending on the instrument employed and on the observer. In the determination of k we must not only take into account that different telescopes have different resolving powers, so that some observers will see more small spots than others will , but we must also consider that the procedures for distinguishing individual spot groups used by different observatories may be different as well. Obviously, the number of groups (g) will have a very high weight in the definition of R.

For his Zurich observations Wolf took $k = 1$. He made his visual estimates using a refractor with an 8-cm objective and a magnifying power of 64. Wolf's observations were continued by Wolfer, his successor at Zurich, and then by Brunner. Wolfer noted that when calculating the relative spot numbers Wolf had omitted pores and had neglected umbral fragmentation

in sunspots following the leading spot, so that his g and f values were both systematically low. To avoid the necessity of modifying the system of Wolf numbers, it was decided in Zurich to introduce a conversion coefficient of 0.6 which would reduce the new Zurich observations to the original Wolf system. At present the Zurich system is accepted as an international standard, and the results of all other observatories are reduced to this system. There is, however, considerable uncertainty involved in reducing R to a single system. The problem is that, as demonstrated by Slonim /22/, the coefficient k does not remain constant even for a given observatory. For example, for the Tashkent Observatory system, which is one of the most stable systems in existence, k varied from 0.6 to 0.9 for different years. This being the situation, we obviously cannot count on a high accuracy for the values of R; the errors in the average values can easily reach 20% or more. Table 3 gives the Wolf numbers R in the Zurich system for over 200 years (1749–1958). The years of maximum are marked with an asterisk.

TABLE 3

Year	R	Year	R	Year	R	Year	R	Year	R	Year	R
1749	80.9	1784	10.2	1819	23.9	1854	20.6	1889	6.3	1924	16.7
1750*	83.4	1785	24.1	1820	15.7	1855	6.7	1890	7.1	1925	44.3
1751	47.7	1786	82.9	1821	6.6	1856	4.3	1891	35.6	1926	63.9
1752	47.8	1787*	132.0	1822	4.0	1857	22.8	1892	73.0	1927	69.0
1753	30.7	1788	130.9	1823	1.8	1858	54.8	1893*	84.9	1928	77.8
1754	12.2	1789	118.1	1824	8.5	1859	93.8	1894	78.0	1929	65.0
1755	9.6	1790	89.9	1825	16.6	1860*	95.7	1895	64.0	1930	35.7
1756	10.2	1791	66.6	1826	36.3	1861	77.2	1896	41.8	1931	21.2
1757	32.4	1792	60.0	1827	49.7	1862	59.1	1897	26.2	1932	11.1
1758	47.6	1793	46.9	1828	62.5	1863	44.0	1898	26.7	1933	5.7
1759	54.0	1794	41.0	1829	67.0	1864	47.0	1899	12.1	1934	8.7
1760	62.9	1795	21.3	1830*	71.0	1865	30.5	1900	9.5	1935	36.1
1761*	85.9	1796	16.0	1831	47.8	1866	16.3	1901	2.7	1936	79.7
1762	61.2	1797	6.4	1832	27.5	1867	7.3	1902	5.0	1937*	114.4
1763	45.1	1798	4.1	1833	8.5	1868	37.3	1903	24.4	1938	109.6
1764	36.4	1799	6.8	1834	13.2	1869	73.9	1904	42.0	1939	88.8
1765	20.9	1800	14.5	1835	56.9	1870*	139.1	1905*	63.5	1940	67.8
1766	11.4	1801	34.0	1836	121.5	1871	111.2	1906	53.8	1941	47.5
1767	37.8	1802	45.0	1837*	138.3	1872	101.7	1907	62.0	1942	30.6
1768	69.8	1803	43.1	1838	103.2	1873	66.3	1908	48.5	1943	16.3
1769*	106.1	1804*	47.5	1839	85.8	1874	44.7	1909	43.9	1944	11.0
1770	100.8	1805	42.2	1840	63.2	1875	17.1	1910	18.6	1945	33.2
1771	81.6	1806	28.1	1841	36.8	1876	11.3	1911	5.7	1946	92.6
1772	66.5	1807	10.1	1842	24.2	1877	12.3	1912	3.6	1947*	151.5
1773	34.8	1808	8.1	1843	10.7	1878	3.4	1913	1.4	1948	136.2
1774	30.6	1809	2.5	1844	15.0	1879	6.0	1914	9.6	1949	134.7
1775	7.0	1810	0.0	1845	40.1	1880	32.3	1915	47.4	1950	83.9
1776	19.8	1811	1.4	1846	61.5	1881	54.3	1916	57.1	1951	69.3
1777	92.5	1812	5.0	1847	98.5	1882	59.7	1917*	103.9	1952	31.5
1778*	154.4	1813	12.2	1848*	124.3	1883*	63.7	1918	80.6	1953	13.9
1779	125.9	1814	13.9	1849	95.9	1884	63.5	1919	63.6	1954	4.4
1780	84.8	1815	35.4	1850	66.5	1885	52.2	1920	37.6	1955	38.0
1781	68.1	1816*	45.8	1851	64.5	1886	25.4	1921	26.1	1956	141.7
1782	38.5	1817	41.1	1852	54.2	1887	13.1	1922	14.2	1957	189.9
1783	22.8	1818	30.4	1853	39.0	1888	6.8	1923	5.8	1958	184.8

The table shows that the length of the period varies over a wide range, from 7 to 17 years. It is thus more correct in this case to substitute the term "cycle" for the term "period."

A solar-activity index which is somewhat more accurate than the Wolf numbers is the average (for some specified period) sunspot area. However,

the statement that this index is more accurate is to a certain extent conditional. During years of sunspot minimum, when the spot area must be estimated according to one or two small sunspots, the error in the daily sunspot areas is almost as high as the relative error in the area of a single sunspot (that is, almost 20%). During years of maximum, however, when the estimate is obtained by adding up the areas of the numerous sunspots on the solar disk, the error of the sum increases approximately in proportion to $\sqrt{n}$, where n is the number of sunspots measured, while the relative error decreases in proportion to $\sqrt{n}$. Therefore, in years of maximum the sunspot area may be determined with a higher accuracy than in years of minimum, and the random error is no more than a few percent.

The average daily spot area is expressed in millionths of the visible surface of the solar hemisphere. The spot areas, as well as the areas occupied by faculae, were previously published in Greenwich Observatory reports. The variation in facula area (upper curve in Figure 20) agrees in general with the variation in spot area, except that in years of minimum it does not dip as low as the area curve. However, since its accuracy is low, the facula area cannot be regarded as a reliable activity index (Figure 20). Starting in 1932, the sunspot areas, as well as the Wolf numbers, have been published in the Proceedings of the Pulkovo Observatory, according to the data of Soviet observatories.

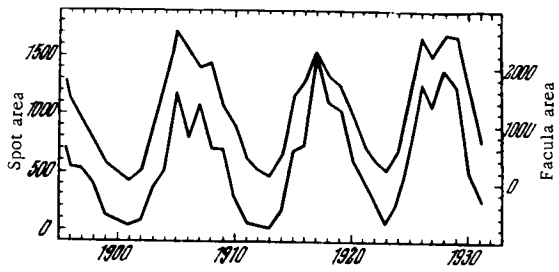


FIGURE 20

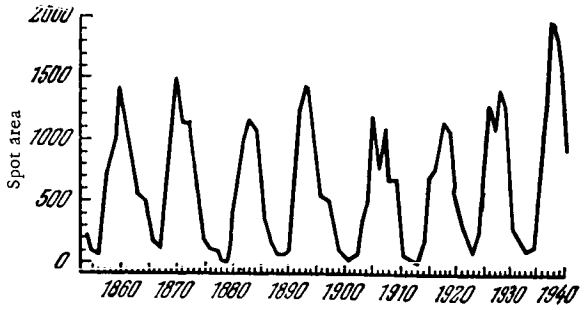


FIGURE 21

From the curves of the Wolf numbers or spot areas (Figure 21) it is easy to see that the different maxima vary in height. The maxima appear to be affected by a second wave of longer period. This long-period variation was first indicated explicitly by Gleissberg. Table 3 shows that high

maxima occurred in 1778, 1837, and 1957. The 1957 - 1958 maximum was higher than the one in 1947, and was actually the highest observed during the last 200 years. The interval between the maxima of 1778 and 1837 is 69 years, while that between 1837 and 1957 is 120 years. Therefore, it is more difficult in this case to reach any conclusion concerning a definite periodicity than it was for the 11-year cycles. The average duration of the long cycle is apparently close to 90 years. On the other hand, as we observed above, the 11-year cycles come in pairs according to the law of magnetic-polarity reversal, so that it would be more accurate to refer to 22-year cycles rather than 11-year cycles. Recently Gnevyshev and Ol' showed that the pairs of 11-year cycles have an almost functional relationship, whereas a relationship with cycles located in the series but not included in the given pair almost does not exist.

Of particular interest is the latitudinal migration of the sunspot zone during the course of the 11-year cycle. At the beginning of a cycle, new sunspots invariably form in both hemispheres at comparatively high latitudes: $30° \leqslant \varphi \leqslant 40°$. Higher than 40°, only small short-lived sunspots are occasionally observed. Subsequently, the sunspot zone begins to drift as a wide belt (about 20° wide) toward the equator, until at the end of the cycle, with the advent of the epoch of minimum, the sunspot zone reaches the equator. Individual spots have been observed almost at the equator. The diagram (Figure 22) constructed by Maunder in 1922 clearly illustrates this phenomenon. In this diagram every spot is plotted against the average date of its observation (the abscissa) as a small streak in the latitude interval (the ordinate) in which it has been observed. The resulting figures resemble the wings of butterflies, so that Maunder's diagram is called a "butterfly" diagram. The latitudinal drift of the sunspot zone was discovered in 1858 by Carrington. Later, Spörer (1894) traced the variation in the spot latitudes back to 1621, so that the variation became known as Spörer's law. Recently, Gnevyshev and Gnevysheva have shown that, although the yearly spot-number variation within a single 11-year cycle differs greatly from the spot-number variation in another 11-year cycle, still the yearly variation in the average spot latitude is the same for all cycles. A definite relationship exists between the variation in spot latitude and the variation in spot number, namely that the greater the number of spots in a cycle is, the higher will be the latitude at which the cycle commences and the higher will be the average latitude of the spots in the year of maximum. In other words, during the strongest cycles (those with the highest spot numbers), the spots start forming at the highest latitudes. Thus, the height of a new cycle can be predicted from the latitude at which the first spots appear.

The asymmetry of the sunspot formation in the northern and southern solar hemispheres is also of interest. The first reliable data on this asymmetry were obtained by Maunder in 1922. Later, this phenomenon was observed for a large number of cycles by Vsekhsvyatskii, and finally, the recent studies of Bezrukova represent a comprehensive analysis of the subject. Bezrukova studied the behavior of the average spot area for the entire time covered by the Greenwich observations. Figure 23 shows the activity curves for the two hemispheres (the solid line is the northern hemisphere and the dashed line is the southern hemisphere). The figure shows that the distribution of spots in the solar hemispheres is not random. In some cycles the sunspot activity appears to be "depressed" in one of the hemispheres.

Many authors have tried to establish an analytical formula describing the 11-year cycle. The most successful in this respect is the study of Stewart and Panofsky, who obtained the following empirical formula for the Wolf number (R):

$$R = F\theta^a e^{-b\theta}, \qquad (21)$$

where F, a, and b are constants, and θ is the number of years which have elapsed from the beginning of the cycle to the epoch in question. The constants F, a, and b vary for different cycles, and Stewart and Panofsky have established that F and b are functions of a. The advantage of formula (21) is its simplicity, but the "eruptive" nature of a spot cycle cannot be established by a formal examination of some purely extrinsic general property of function (21) and the actual functions describing eruptive processes.

Although the latitudinal position of the sunspot zone is specified by Spörer's law, the longitudinal distribution of spots is not far from random. Nevertheless, many observers have noted that spot groups repeatedly appear at almost the same longitude. There appears to be some kind of active longitude within which the solar-activity center remains for several months, and sometimes for over a year. In time this longitude becomes inactive, and the activity centers shift to other longitudes. These facts, as we have already observed, are in good agreement with present-day conceptions of the stability of sunspot magnetic fields. The spots do not disintegrate, but rather appear to just "sink" into the photosphere, and thus may emerge in the same place on some later occasion.

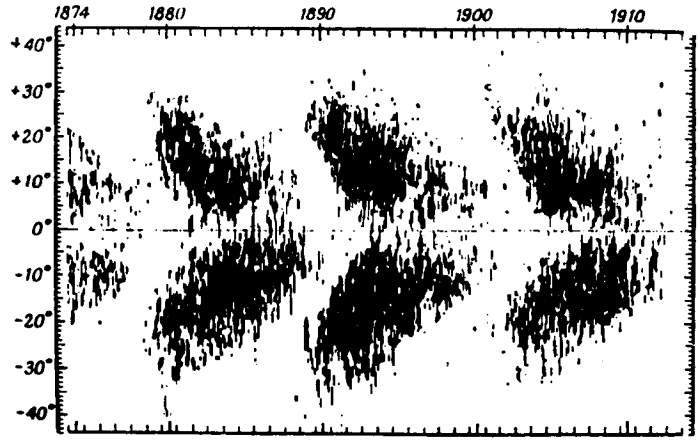


FIGURE 22

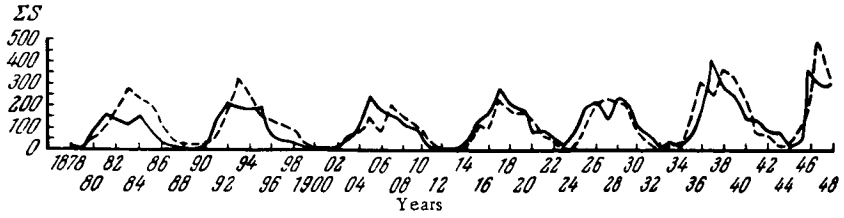


FIGURE 23

§19. The cyclic nature of sunspot formation. The regular reversal of the magnetic polarities of sunspot groups, the regular latitudinal drift of sunspot zones, and the 11-year cycles governing all surface phenomena on the sun all seem to indicate that a single mechanism is responsible for solar activity.

The first to attempt to understand how this mechanism operates was Bjerknes (1926), who observed that nonuniform rotation of the sun is possible only in the presence of a general circulation of gas masses in the solar interior. Gas streams should move toward the poles at great depths below the photosphere, and then should return to the equator near (or actually at) the surface of the photosphere (Figure 24). Such a circulation could even be caused by the solar rotation. Actually, at about the same time, von Zeipel and later Eddington /20/ showed that, given a radial energy transfer from the center of the sun, the temperature at an equipotential surface of a rotating star (in this case the surface of the photosphere can be regarded as such) should be proportional to $\sqrt[4]{g}$ (where g is the acceleration of gravity). In other words, the temperature at the solar poles should be higher than that at the equator. Later, V. A. Krat /23/ demonstrated that the Bjerknes convective flux is stationary, that is, that it cannot be destroyed by viscosity forces. Finally, Rubashev has recently proven that the circulation currents on the sun should split in the longitudinal direction, thereby forming alternating regions of high and low pressure at the surface of the photosphere. According to Bjerknes, sunspots are vortex tubes which are transferred by the currents of the general circulation. Since near the photosphere the circulation is directed from high latitudes toward the equator, Bjerknes assumed that this explains Spörer's law of latitudes. The 11-year periodicity of sunspot formation, however, cannot be explained by the Bjerknes theory. In fact, all the foregoing ideas were created by analogy with various terrestrial phenomena. The existence of a general circulation of gas masses in the earth's atmosphere led some authors to attempt to construct a hydrodynamic theory of circulation for the sun. However, such a theory must of necessity be incomplete, and even to some extent fallacious, since solar phenomena are shaped to a great degree by electrodynamic forces, and, in particular, by the forces of induced magnetic fields, which are apparently irregular in nature. These magnetic forces should act primarily as ponderomotive forces resisting the motion of the gas currents. For example, if a sufficiently strong dipole magnetic field were set up on the sun, then circulation would be altogether impossible. Currents intersecting the magnetic lines of force of the dipole would either stop due to the field resistance or else would deform the field, transforming it from a dipole field into some very complicated twisted pattern of lines of force. In the latter case the field would be transported by the circulation currents.

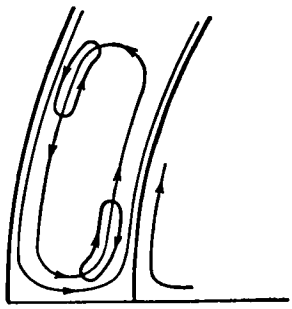


FIGURE 24

No satisfactory magnetohydrodynamic theory has been developed so far to explain the cyclic nature of solar activity. Many recent theories (those of Alfvén, Walén, and others) have proved to be inconsistent /20/.

REFERENCES FOR CHAPTER III

1. JANSSEN, J. — Ann. Obs. Paris, Meudon, Vol. 1: 103. 1896.
2. GANSKII, A.P. — Izvestiya Glavnoi astronomicheskoi observatorii, No. 25. 1908; Trudy Pulkovoskoi observatorii, Vol. 1: 81. 1905 and Vol. 3: 1. 1918.
3. KRAT, V.A. — Izvestiya Glavnoi astronomicheskoi observatorii, No. 152. 1954; KRAT, V.A. and N.M. GOL'DBERG. — Izvestiya Glavnoi astronomicheskoi observatorii, No. 155. 1956.
4. UNSÖLD, A. Physics of Stellar Atmospheres. — Russian Translation from German, Edited by E.R. Mustel', Izdatel'stvo inostrannoi literatury, 1949.
5. NEWTON, H.W. and M.L. NUNN. — M.N., Vol. 111: 413. 1951.
6. KRAT, T.V. — Izvestiya Glavnoi astronomicheskoi observatorii, No. 137. 1947 and No. 140. 1947.
7. TEN BRUGGENCATE, P. and H. VON KLÜBER. — Zs. f. Ap, Vol. 18: 284. 1939.
8. ABHYANKAR, K.D. and A.S. RAMANATHAN. — Ap.J., Vol. 121: 750. 1955.
9. AMBARTSUMYAN, V.A. et al. Teoreticheskaya astrofizika (Theoretical Astrophysics). — Gostekhizdat. 1952 [English Translation. Pergamon].
10. KRAT, T.V. — Astronomicheskii zhurnal, Vol. 24: 329. 1947.
11. MITROPOL'SKAYA, O.N. — Izvestiya Krymskoi astrofizicheskoi observatorii, Vol. 8: 93. 1952; Vol. 11: 152. 1954; Vol. 15: 130. 1955.
12. MEL'NIKOV, O.A. and S.S. ZHURAVLEV. — Vestnik Leningradskogo universiteta, No. 13, Issue 3. 1956.
13. MEL'NIKOV, O.A. and S.S. ZHURAVLEV. — Vestnik Leningradskogo universiteta, No. 8. 1955.
14. HALE, G.E. and S.B. NICHOLSON. Magnetic Observations of Sunspots, 1917—1924 (Washington, Carnegie Institution of Washington). Part I, p. 56.
15. BABCOCK, H.W. and H.D. BABCOCK. [Mapping the Magnetic Fields of the Sun]. — In Chapter 9 of "The Sun", edited by G. Kuiper, Russian Translation from English, Izdatel'stvo inostrannoi literatury. 1956.
16. NIKULIN, I.S., A.B. SEVERNYI, and V.E. STEPANOV. — Izvestiya Krymskoi astrofizicheskoi observatorii, Vol. 19: 3. 1958.
17. KOTLYAR, L.M. — Izvestiya Glavnoi astronomicheskoi observatorii, No. 163. 1959.
18. STEPANOV, V.E. — Uchenye zapiski L'vovskogo gosudarstvennogo universiteta, Vol. 15, seriya fiziko-matematicheskaya, No. 4. 1949.
19. BIERMANN, L. — Vierteljahrsschr. Astr. Ges., Vol. 76: 194. 1941.
20. GUREVICH, L. and A. LEBEDINSKII. — Zhurnal eksperimental'noi i teoreticheskoi fiziki, Vol. 16: 832 and 840. 1946.
21. COWLING, T.G. [Solar Electrodynamics]. — Chapter 8 of "The Sun", Edited by G. Kuiper, Russian Translation from English, Izdatel'stvo inostrannoi literatury. 1956.
22. SLONIM, Yu. M. — Trudy Tashkentskoi astronomicheskoi observatorii, Seriya II, Vol. 1: 47. 1950.
23. KRAT, V.A. Figury ravnovesiya nebesnykh tel (Equilibrium Figures of Celestial Bodies). — Gostekhizdat. 1950.

V. P. VYAZANITSYN and V. M. SOBOLEV

N67-15488

Chapter IV

THE CHROMOSPHERE

§20. The flash spectrum. General properties of the chromosphere. Just prior to the onset of a total solar eclipse it is possible to observe, by placing a spectrograph slit tangent to the solar limb, that after the gradual attenuation of the solar spectrum the dark Fraunhofer lines are abruptly transformed into bright lines. The corresponding emission spectrum, which is observed for about 1.5 seconds near the 2nd eclipse contact (and analogously near the 3rd contact, but in reverse order), is called a flash spectrum. The abrupt reversal at the onset of the total phase is actually observed for weak Fraunhofer lines only, whereas for stronger lines the reversal is not instantaneous.

The emission spectrum at the solar limb pertains to a rarefied gas layer $\sim 10''$ high, located adjacent to the photosphere. This layer is known as the chromosphere (Lockyer), because of the bright-red color it presents to the naked eye. The flash spectrum actually originates in the lower part of the chromosphere or, to use the older terminology, in the reversing layer. The latter term came about due to the initial conception of a flash spectrum as an exact reversal of the Fraunhofer spectrum, which was assumed to originate in this layer. Since its discovery in 1870 the flash spectrum has been an inseparable part of the program of every eclipse expedition, and it has been studied quite intensively. The exact wavelengths of the chromospheric lines have been measured, their intensities estimated, and the lines identified; in addition, the observable spectral region has been extended in both directions and the observational techniques have been perfected.

Total solar eclipses provide the most favorable conditions for observations of the chromosphere, which are otherwise severely restricted by the masking effects of the bright solar light scattered in the earth's atmosphere and in the instrument. Consequently, eclipse observations still represent the main source of information on the physical properties of the chromosphere.

On the basis of several instantaneous photographs taken with prism cameras during the 1898 eclipse, Fowler and Lockyer first demonstrated that the flash spectrum does not constitute an exact reversal of the Fraunhofer spectrum. This was evident from the presence of helium lines, from the higher number (compared to the Fraunhofer spectrum) of Balmer hydrogen lines, and from the increased intensity of spark lines (which were the most numerous for Ti and Fe and fewer for Cr, V, and I) relative to arc lines. Fowler and Lockyer measured a total of 856 lines, most of which could be identified. They noted the great heights into the chromosphere of H, He, Ca, and also of the line λ 4686, later identified by Fowler as He^+.

During the same eclipse, Evershed traced out the spectrum in the ultraviolet region to λ = 3342 Å; he obtained 30 Balmer lines and discovered the continuous spectrum of the chromosphere and the prominences. Evershed decided that the flash spectrum originates in the upper part of

the gas layer responsible for the Fraunhofer spectrum. In 1906 Dyson published the results of a study of slit spectrograms taken during the eclipses of 1900, 1901, and 1905. About 1200 lines were recorded in the λλ 3300—5875 region of the flash spectrum, and a relative strengthening of the helium lines in the upper part of the chromosphere was observed.

The increased intensity of spark lines relative to arc lines in the chromospheric spectrum was universally acknowledged after the 1905 eclipse, in particular due to a more accurate determination of the line wavelengths. The interpretation of this phenomenon, however, met with considerable difficulties. Since spark lines (which are produced by ionized, as distinct from neutral, metals) are assumed to correspond to a higher temperature, it appeared that the chromosphere was hotter than the underlying layers, which was considered to be an absurd result. This problem was solved only with the development, by Saha (1920), of the theory of ionization equilibrium, according to which the degree of ionization of a gas is determined by pressure (more precisely, by the partial electron pressure), as well as by temperature. The high ionization could thus be attributed to a low pressure in the chromosphere. Recently, however, with the discovery of a temperature inversion somewhere between the photosphere and the corona, the situation has become somewhat more complicated. The ionization of metals observed in the flash spectrum, though, no longer requires higher temperatures than those in the photosphere and is fully explainable in terms of a lower pressure.

During the 1926 eclipse Davidson and Stratton used a quartz slit spectrograph to photograph the λλ 3066—4216 region of the chromospheric spectrum. The line wavelengths were measured and most of the lines were identified. The resulting line catalog also listed line-intensity estimates on an arbitrary scale for the flash spectrum and for a high prominence. The Balmer hydrogen lines were resolved up to H_{36}, after which they became a continuum. Photographs taken with two prism cameras were used to derive, from the lengths of the chromospheric arcs, the "heights" of the lines in the λλ 3820—6678 region; the overall level of most of the weaker lines was found to be 400 to 500 km, while for many of the lines it was < 400 km. On the whole, the longest arcs correspond to the strongest lines, helium constituting an exception since its arcs are long but quite weak. During the same eclipse Curtis and Burns photographed the infrared chromospheric spectrum; they distinguished the O triplet λλ 7772, 7774, 7775, and the Ca^+ lines λλ 8542 and 8662.

One of the most important general studies of the flash spectrum was that carried out by Menzel, on the basis of the spectrograms taken by Campbell during the eclipses of 1900, 1905, and 1908 /1/. After processing a large amount of observational material, Menzel compiled a catalog of about 4000 lines of the flash spectrum in the λλ 3229—5328 range. This catalog included the following basic data: the line wavelengths, the line-intensity estimates on the Rowland scale for the base of the chromosphere (the point at which the background continuum disappears from photographs), and the "heights" of the lines (from the base of the chromosphere). These basic data were also supplemented by several other items of information (depths of the line-reversal point below the chromosphere base, identification of lines, etc.).

The excellent catalog of flash-spectrum lines compiled by Mitchell from his own observations during several solar eclipses /2/ has been exceptionally

important not only for solar-atmosphere research but also for astrophysics as a whole. Mitchell's catalog was first published in 1913, on the basis of photographs of the 1905 eclipse. Subsequently, as new data became available (the other spectrograms of Mitchell for 1905 and his spectrograms in 1925, 1930, and 1937, as well as, in part, the observations of Davidson and Stratton in 1926) and the observational techniques became perfected, the catalog was enlarged and completely revised twice (in 1930 and 1947), in order to ensure the accuracy of the data and the uniformity of the scales used. The main apparatus used by Mitchell was a slitless spectrograph with a parabolic grating, the spectra being photographed onto celluloid films. In the latest edition (1947) the catalog included 3500 lines in the $\lambda\lambda$ 3066—8863 region.* The basic data listed in the catalog are the line-intensity estimates for the flash spectrum on the Rowland scale and the "heights" of the lines as determined from the lengths of chromospheric arcs on the photographs. These data in Mitchell's catalog represent the "weighted means" of four eclipses, where the averaging was made over conditions which depended on the photographic plate or on the particular eclipse.

The catalogs of Menzel and Mitchell include a very comprehensive body of observational data derived from general studies of the flash spectrum. These data represent a valuable addition to the information on the solar spectrum contained in the Revised Rowland. They have also served as a basis for various theoretical investigations and have led to many important conclusions concerning the chromosphere. For example, using the line-intensity estimates, Menzel, and also Mitchell and Williams, were the first to make a general quantitative chemical analysis of the chromosphere (see §3), and from the "heights" of the lines they determined the density-distribution gradients of various elements over the height of the chromosphere. It should be noted that, although at present no real significance is attached to the observed line heights, still Wildt in a more recent work /3/ has determined the composition of the chromosphere and has derived the density gradients of the elements exclusively from the data for the "heights" of the lines.

Table 1 lists the brightest lines in the spectrum of the chromosphere, according to Mitchell /2/. Here $I_\odot$ is the Rowland estimate of the corresponding line intensity for the solar spectrum, I_c is the intensity estimate of the chromospheric line on the same scale, H is the "height" of the line, and E.P. is the excitation potential of the line in electron volts.

We see from Table 1 that on the whole the brighter lines have greater "heights." This is consistent with the assumption that the emission gradients of lines of the different elements do not differ much from one another, since the observed line "heights" actually represent some minimum emission intensity for the chromosphere, as specified by the sensitivity threshold of the photographic plate in the given spectral region. A typical example of the decrease in line "height" with the decrease in line intensity is provided, in particular, by the Balmer series of hydrogen. Helium lines, on the other hand, constitute a noticeable exception, since they reach much greater "heights" than equally bright lines of other elements. The brightest lines in the chromospheric spectrum are the lines of ionized calcium, the Balmer lines of hydrogen, and the D_3 line (λ5875.6) of helium. Accordingly, the greatest "heights" are attained by calcium lines (14,000 km) hydrogen lines (13,000 km) and helium lines (7500 km).

* The Revised Rowland, by the way, starts at λ=2975 Å.

TABLE 1

λ	Element	$I_\odot$	I_c	H, km	E.P.	λ	Element	$I_\odot$	I_c	H, km	E.P.	λ	Element	$I_\odot$	I_c	H, km	E.P.
3234.49	Ti^+	10	30	2000	3.86	3829.35	Mg	10	40	5000	5.92	4555.89	Fe^+	3	25	1200	5.52
3236.59	Ti^+	7	25	2000	3.84	3832.34	Mg	15	50	5000	5.92	4563.76	Ti^+	4	30	2500	3.92
3242.00	Ti^+	7	25	2000	3.81	3835.54	H_9	—	120	7000	13.37	4572.00	Ti^+	6	40	2500	4.26
3349.41	Ti^+	9	40	2500	3.73	3838.30	Mg	25	60	6000	5.92	4583.86	Fe^+	4	30	1800	5.49
3372.84	Ti^+	10	40	2500	3.67	3859.87	Fe	20	40	2500	3.20	4685.83	He^+	—	2N	3500	50.80
3408.81	Cr^+	3	25	1500	6.09	3889.20	H_ζ	—	140	8000	13.33	4713.15	He	—	15	5000	23.49
3421.25	Cr^+	4	25	1500	6.02	3900.54	Ti^+	5	35	2000	4.29	4861.50	H_β	30	160	9000	12.69
3441.97	Mn^+	6	25	1500	5.35	3933.90	Ca^+	1000	200	14000	3.14	4922.01	He	—	6	3000	23.63
3460.32	Mn^+	4	25	1500	5.37	3944.03	Al	15	25	2000	3.13	4923.96	Fe	5	35	2000	5.39
3474.11	Mn^+	4	25	1500	5.35	3961.51	Al	20	35	2000	3.13	4934.08	Ba^+	7	30	1800	2.50
3535.41	Ti^+	4	25	1200	5.54	3968.70	Ca^+	700	180	14000	3.11	5015.68	He	—	8	3500	22.99
3581.22	Fe	30	25	1500	4.30	3970.25	H_ε	5	120	8000	13.26	5018.44	Fe^+	4	40	2000	5.34
3613.82	Sc^+	7	30	1500	3.44	4026.28	He	—	35	6000	23.94	5172.65	Mg	20	60	3000	5.09
3641.35	Ti^+	4	25	1500	4.62	4030.79	Mn	9	25	1500	3.06	5183.58	Mg	30	70	3500	5.09
3679.34	H_{21}	—	30	2500	13.51	4045.84	Fe	30	40	2000	4.53	5208.37	Cr	5	25	1200	3.31
3682.82	H_{20}	—	30	2500	13.51	4063.62	Fe	20	25	1800	4.59	5316.67	Fe^+	6	40	1200	5.46
3685.25	Ti^+	10	90	6000	3.92	4071.75	Fe	15	25	1500	4.63	5875.64	$He(D_3)$	—	100	7500	22.97
3686.83	H_{19}	—	35	3000	13.50	4077.83	Sr^+	8	80	6000	3.03	5889.09	$Na(D_2)$	30	40	1500	2.10
3691.62	H_{18}	—	40	3500	13.50	4101.85	H_δ	40	140	8000	13.16	5895.99	$Na(D_1)$	20	35	1500	2.09
3697.21	H_{17}	—	40	4000	13.49	4215.70	$Sr^+ - CN$	5	60	6000	2.93	6162.19	Ca	15	25	1000	3.89
3703.89	H_{16}	—	45	4500	13.49	4226.74	Ca	20	40	4000	2.92	6496.88	Ba^+	4	30	1500	2.50
3712.06	H_{15}	—	50	5000	13.48	4233.22	Fe^+	4	30	2200	5.49	6562.80	H_α	40	200	12000	12.04
3719.94	Fe	40	35	2000	3.32	4246.90	Sc^+	5	50	5000	3.22	6678.10	He — Fe	8	30	2200	22.97
3722.00	H_{14}	—	60	5500	13.47	4254.36	Cr	8	35	2000	2.90	7065.18	He	—	60	7500	22.62
3734.45	H_{13}	—	70	6000	13.46	4290.18	Ti^+	2	25	2000	4.04	7771.95	O	8	45	6000	10.69
3741.64	Ti^+	4	30	2000	4.87	4300.05	Ti^+	3	35	2500	4.05	7774.18	O	6	30	6000	10.69
3745.78	Fe	14	30	2000	3.38	4340.63	H_γ	20	150	8000	13.00	7775.39	O	5	25	6000	10.69
3750.25	H_{12}	—	75	6000	13.45	4383.54	Fe	15	30	2000	4.29	8446.33	O	—	25	3000	10.94
3759.33	Ti^+	12	65	6000	3.89	4404.79	Fe	10	25	1800	4.35	8446.76	O	—			10.94
3761.33	Ti^+	7	75	6000	3.85	4443.85	Ti^+	5	35	3000	3.85	8498.06	Ca^+	—	140	10000	3.14
3761.88	Ti^+	3			5.86	4468.48	Ti^+	5	40	2500	3.89	8542.14	Ca^+	—	160	12000	3.11
3770.72	H_{11}	—	80	6000	13.43	4471.54	He	—	90	7500	23.63	8598.40	H	—	25	6000	13.47
3798.02	H_{10}	—	100	6000	13.40	4501.28	Ti^+	5	35	2500	3.85	8662.17	Ca^+	—	180	12000	3.14
3820.43	Fe	25	25	1600	4.09	4554.11	Ba^+	8	50	2500	2.71	8750.47	H	—	25	6500	13.45

A comparison of the line intensities for the solar [photospheric] spectrum and the flash spectrum shows that: 1) the relative line intensities do not correspond to one another for the two cases; 2) the intensities of ionized lines relative to those of neutral lines are higher in the flash spectrum than in the solar spectrum; 3) no helium lines are observed in the solar spectrum (not even the bright D_3 helium line); and 4) the number of hydrogen lines in the solar spectrum is much lower than in the flash spectrum.

One of the basic problems of the theory of the chromosphere is how to explain the fact that, in spite of the comparatively low overall level of excitation for atoms in the chromosphere (E. P. $\approx$ 3 to 5 eV), the flash spectrum contains bright lines of hydrogen (E. P. $\approx$ 13.5 eV) and helium (E. P. $\approx$ 23.5 eV), and even contains the He^+ line 4686 Å (E. P. = 50.80 eV).

§21. Spectrophotometric measurements and chromospheric-line profiles. Estimates of line intensities are now of secondary importance. Beginning in the 1920's, the method of exact spectrophotometry acquired a basic and ever-increasing significance, since it makes it possible to obtain more reliable and detailed data on the physical state of the chromosphere. However, because of the complexity of the technique and for certain other reasons as well, only relatively few trustworthy spectrophotometric measurements of the chromosphere have been obtained.

The first spectrophotometric study of the chromosphere, the results of which until recently were primarily used in all theoretical computations, was carried out by Cillié and Menzel, using the flash-spectrum photographs obtained during the 1932 eclipse* /4/. The observations were made using two spectrographs, the photographs on both spectrographs being taken simultaneously. The measurements gave absolute line intensities in the λλ 5876 – 3323 range, for successive heights of the lunar limb (not above 2570 km), for two regions in the lower chromosphere (the normal and high-excitation regions). The intensities were expressed as the radiation (in ergs per second) emitted in all directions by a strip of chromosphere 1 cm wide, perpendicular to the solar limb. Some values were also obtained for the intensities of the Balmer continuum at λλ 3640 and 3500.

On the basis of their observations, Cillié and Menzel calculated several important characteristics of the chromosphere. For most of the lines measured, the emission gradients of the normal chromosphere were derived from the intensities for lunar-limb heights of 670 and 1500 km. The electron density at the base of the chromosphere was found from the intensity of the Balmer continuum (see §7). By applying the theory of growth curves to take into account the self-absorption in the H and K lines of Ca II and in the first lines of the Balmer series for hydrogen, Cillié and Menzel established the densities at the base of the chromosphere and the concentration gradients of the corresponding atoms (see §5). A comparison of the observed and theoretical line intensities for multiplets confirmed the conclusion of Pannekoek and Minnaert /5/ that self-absorption occurs for metals in the lower chromosphere. Estimates were made of the concentration gradients of some metals, according to their emission gradients, by means of determining the differences in the heights at which the components of the corresponding multiplets have equal intensities. It was found that for all elements the concentration gradient for neutral atoms is greater than that for ionized atoms, a situation which is clearly due to an increase in ionization with height in the chromosphere.

* In their discussion the authors also refer in part to observations made during the 1930 eclipse.

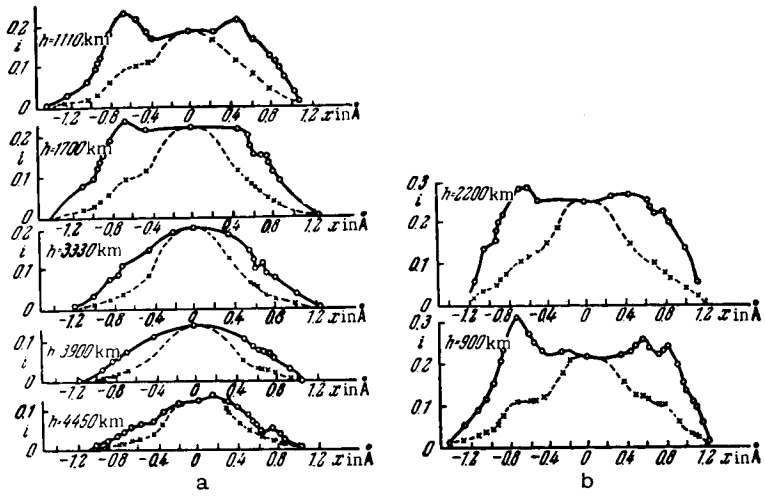


FIGURE 25

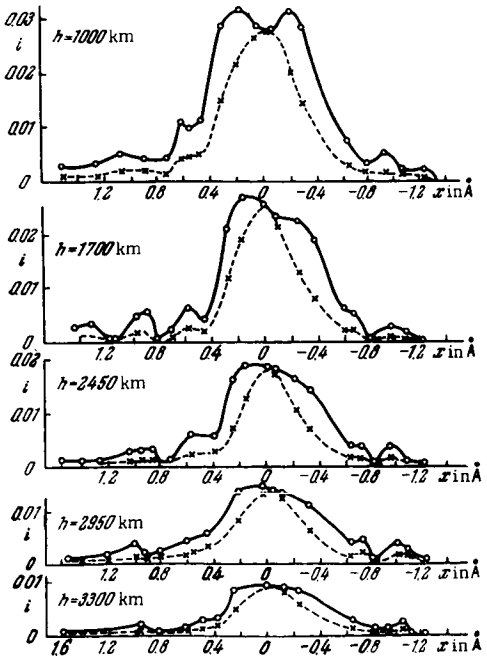


FIGURE 26

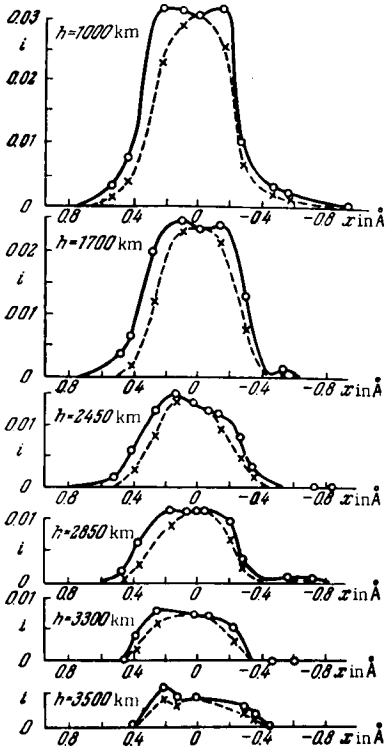


FIGURE 27

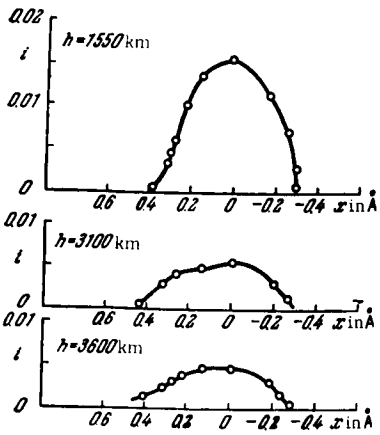


FIGURE 28

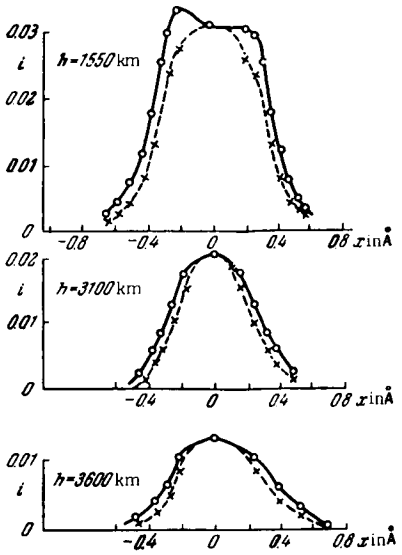


FIGURE 29a

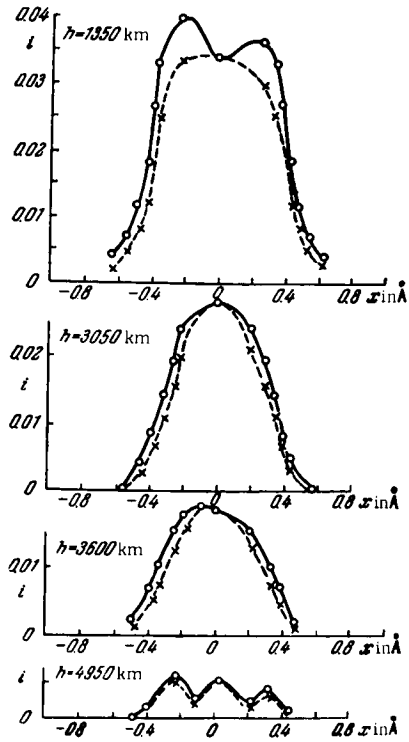


FIGURE 29b

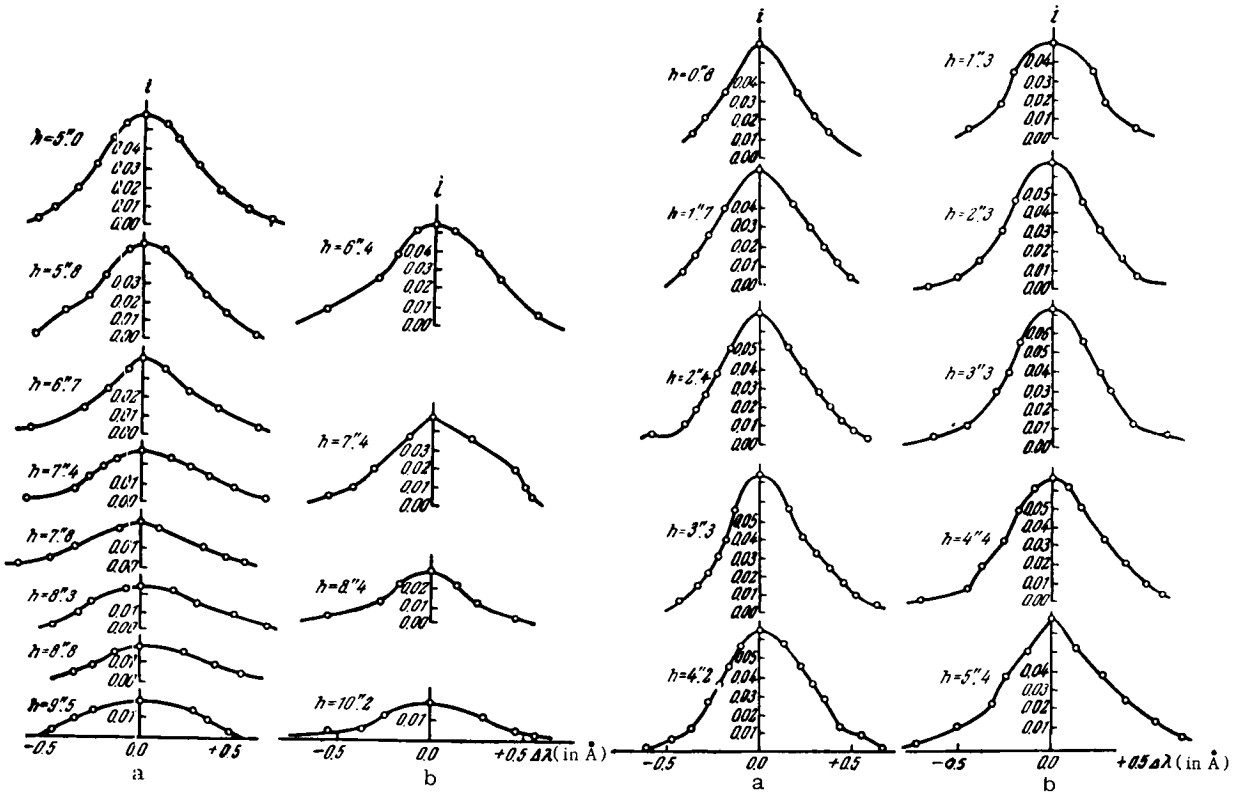

i
h=5".0
h=5".8
h=6".7
h=7".4
h=7".8
h=8".3
h=8".8
h=9".5
h=6".4
h=7".4
h=8".4
h=10".2
-0.5
0.0
+0.5
+0.5 Δλ(in Å)
a
b
h=0".8
h=1".7
h=2".4
h=3".3
h=4".2
h=1".3
h=2".3
h=3".3
h=4".4
h=5".4

FIGURE 30

Spectrophotometric studies of the chromosphere were carried out during more recent solar eclipses by Vyazanitsyn /6/, V. A. Krat /7/, Lindblad and Kristensen /8/, Athay, Billings, Evans, and Roberts /9/, Houtgast and Zwaan /10/, and others. These studies have widened the observable spectral region and the range of heights in the photosphere and have made it possible to improve the results of previous measurements (of the line intensities, the gradients, etc.).

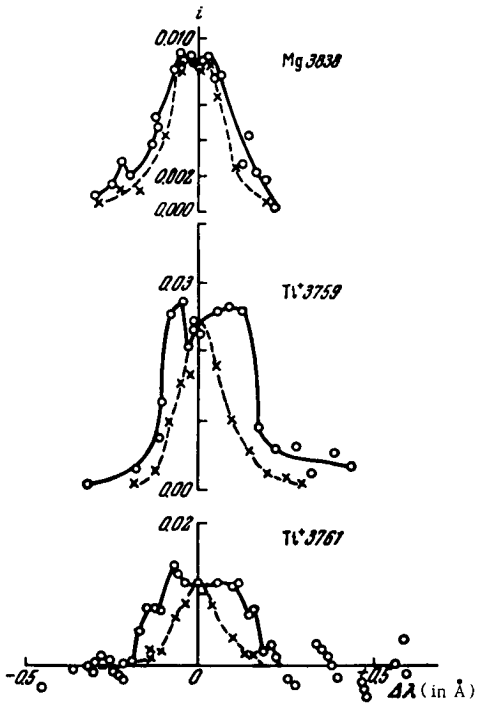


FIGURE 31

The most reliable data on the line profiles are at present obtained during extra-eclipse observations utilizing high-dispersion spectrographs. Such observations were made at Pulkovo from 1953 to 1959 with a horizontal solar telescope giving a solar image 160 mm in diameter /11, 12/. The dispersion of the spectrograph was about 1 Å /mm in the second order. The spectrograms were carefully processed, and the effects which scattered sky light and image vibration have on the line profile were taken into consideration. Some profiles of chromospheric lines are given in Figures 25 through 31 (Figure 25 shows the H_x profiles at various heights, while Figures 26, 27, 28, 29a, 29b, and 30 show the profiles of H_γ, H_δ, H_ε, $H(Ca^+)$, $K(Ca^+)$, and D_3 respectively. The D_3 profiles were obtained from two spectrograms for 10 May 1954. Figure 31 shows profiles of Mg λ 3838, Ti^+ λ 3759 and λ 3761. Circles indicate observed data and crosses indicate reduced values). The curves show that the line profiles of hydrogen, ionized calcium, and magnesium at low heights are double-peaked. This is apparently the case because the emission in these lines is produced by

the scattering of photospheric radiation. Therefore, the radiant flux at the wings of the profile will be greater than at the center. The profile of a chromospheric line in part reflects the profile of the corresponding Fraunhofer line. Consequently, the intensity at some point on the chromospheric-line profile a distance x from the line center will, to a first approximation, be proportional to the intensity of the Fraunhofer line at a point which is also a distance x from the line center. After these effects are eliminated, "rectified" line profiles will be obtained which are single-peaked, in contrast to the observed double-peaked profiles. There are also some other factors which cause the double-peaked profiles. Light-scattering effects result in a gradual reduction of the radiation density in the medium, particularly at points of high optical thickness (at the center of a strong line). As a result, the radiation deficit at the line center may produce some depression of the profile.

The principal conclusion to be drawn from measurements of the profiles of chromospheric lines concerns the presence of random motion ("turbulence") in the chromosphere. This was first established by Unsöld, from the profiles of the H and K lines of Ca II and the D_3 line of He /13/, and it was confirmed by later observations, although different observers have obtained different mean turbulent velocities /1, 14, 15/. It can be shown that constancy of the quantities $\Delta\lambda_D \frac{4000\text{Å}}{\lambda}$ is possible only when the mean turbulent velocity is virtually independent of the height h and when the temperature of the chromosphere is so low that thermal expansion is indistinguishable from expansion due to the radial velocities in the gas. If we assume that $\Delta\lambda_0 \frac{4000\text{Å}}{\lambda}$ for hydrogen has an average value of 0.29 Å, and that for Ca II it is equal to 0.30 Å, then we obtain mean turbulent velocities which are almost equal (about 22 km/sec).

§22. **The chemical composition of the chromosphere.** The quantitative composition of the chromosphere was first determined by Menzel in 1931 from Lick Observatory eclipse spectrograms /1/, and then somewhat later by Mitchell and Williams on the basis of Mitchell's 1930 flash-spectrum catalog /16/. In both studies Russell's method was applied (see §10), but in this case to emission lines.

The values obtained for $\lg A$, where A indicates the relative abundances (numbers of atoms) of the elements, are listed in Table 2. This table also gives the values of $\lg A$ for some of the most abundant elements in the reversing layer (that is, in the photosphere), expressed on the original Russell scale and representing averages of the determinations of Russell, Menzel, Goldberg, Unsöld, Hunaerts, and Claas /17/.

A comparison of the results for the chromosphere with those for the reversing layer indicates that, to a first approximation, the relative abundances of the elements are the same in both cases. Nevertheless, it has been established that hydrogen and the rare earths (not shown in Table 2) are at least 10 times (and according to Menzel hydrogen is 100 times) more abundant in the chromosphere than in the reversing layer, and that the chromosphere is also richer in Sc, Ti, and Cr but is poorer in Fe, Co, and Ca. The abundances of the rare earths are especially inaccurate, due to the lack of suitable spectroscopic data and due to the weakness of the corresponding lines in the flash spectrum. On the other hand, in addition

to the calibration errors, the relative abundances depend strongly on the corrections for deviations from thermodynamic equilibrium, which are determined empirically; in practice, moreover, it is very difficult to distinguish between the two sources of error. This factor is particularly significant for elements with high excitation potentials, such as hydrogen and helium. According to Menzel, He I is at least 10^6 times, and He II is 10^{25} times, more abundant than the values given by thermodynamic considerations.

TABLE 2

Element	Photo-sphere	Chromosphere			Corona*	Sun
		Menzel	Mitchell and Williams	Wildt		
1 H	12.0	13.6	—	11.7	11.7	12.0
2 He	11.3	—	—	—	—	11.3
3 C	8.1	—	—	—	—	8.1
4 N	8.7	—	—	—	—	8.7
5 O	8.9	—	—	(9.2	—	8.9
11 Na	6.5	7.2	6.3	6.4	—	6.5
12 Mg	7.8	7.4	7.6	7.8	—	7.8
13 Al	6.4	6.2	—	6.3	—	6.3
14 Si	7.4	5.7	7.7	—	—	7.5
16 S	6.8	—	—	—	—	6.8
19 K	5.5	—	—	—	—	5.6
20 Ca	6.5	6.6	6.0	6.5	6.0	6.4
21 Sc	3.4	3.8	4.1	—	—	3.6
22 Ti	5.0	5.7	5.4	—	—	5.1
23 V	4.4	5.6	4.7	—	—	4.4
24 Cr	5.5	6.1	6.0	—	—	5.6
25 Mn	5.5	6.4	5.6	—	—	5.6
26 Fe	7.3	7.0	6.5	7.2	7.5	7.2
27 Co	5.8	5.0	4.6	—	—	5.0
28 Ni	5.8	6.0	4.9	—	6.1	5.7
29 Cu	4.6	—	3.5	3.7	—	4.4
30 Zn	5.0	4.3	4.9	—	—	5.1
38 Sr	3.2	3.1	3.9	3.6	—	3.4
39 Y	3.0	3.5	3.7	—	—	3.2
40 Zr	2.4	3.1	3.8	—	—	2.4
56 Ba	2.9	3.7	3.8	—	—	3.3

* The hydrogen abundance is reduced to the average scale for the other three elements.

Wildt /3/ studied the chemical composition of the chromosphere on the basis of Mitchell's 1947 catalog; for this purpose Wildt used the line "heights" rather than the total line intensities, since the latter cannot be rid of the influence of the density gradients.

If we compare the various results for the abundances of elements in the chromosphere, then we see that the values are not the same in every particular case but that on the whole the discrepancies are of the same order of magnitude as the differences between the abundances in the chromosphere and in the reversing layer. This suggests the highly probable conclusion that the reversing layer and chromosphere have identical chemical compositions. Table 2 also lists the relative abundances of H, Ca, Fe, and Ni observed in the solar corona, according to the calculations of Woolley and Allen /18/. This result was also obtained previously by Shklovskii. The relative abundances in the corona appear to be approxi - mately the same as those in the chromosphere and reversing layer. Taking into account the great difference in the physical conditions in the photosphere,

chromosphere, and corona, we reach the very definite conclusion that the chemical composition of the solar atmosphere is identical for all levels. This composition is apparently best represented by the values in the last column of Table 2, which were obtained by averaging the most reliable results for the reversing layer and chromosphere according to Mitchell /17/.

The special features of the general composition of the reversing layer or photosphere which were discussed previously (§ 10) thus also apply to the chromosphere. The most important factor in this respect is the very great predominance in the solar atmosphere of the element hydrogen, the atoms of which comprise 83.2% of all the atoms there (and together with helium comprise 99.87% of all the atoms).

§ 23. The density distribution with height in the chromosphere. Measurements of the chromospheric-line intensities at various distances from the solar limb make it possible to establish in each case the variation in the density (concentration) of the emitting atoms with height in the chromosphere, provided the chromosphere is once again regarded as a homogeneous layer.

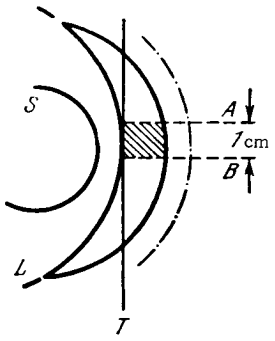


FIGURE 32

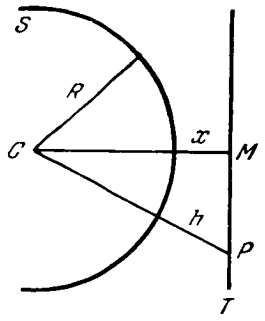


FIGURE 33

Since the resolving powers of the instruments used at present are insufficient, direct measurements of the chromospheric-line intensities at various distances from the solar limb, using slit or slitless stationary spectrographs, are very difficult. For such measurements, several successive spectrograms are generally used, taken near the 2nd or 3rd eclipse contact with a slitless spectrograph. The intensities of the lines (the chromospheric arcs) measured on these spectrograms represent the total emission of the part of the chromosphere which, at the given point on the solar limb and at the given eclipse moment, protrudes past the lunar limb. Accordingly, the line-intensity measurements are related to the volume of the chromosphere bounded by three planes parallel to the line of sight (see Figure 32): the plane T tangent to the lunar limb L, and the two parallel planes A and B separated 1 cm from one another and oriented perpendicular to the solar limb S at the given point. Let us denote $E(x)$ as the emission in a given spectral line within the specified volume of the chromosphere in a solid angle $\omega = 1$, where x is the distance of the lunar limb from the solar limb at the point in question and at the given moment

of the eclipse. The derivative of this function is the emission intensity of the chromosphere at the corresponding height:

$$-\frac{dE(x)}{dx} = I(x). \tag{1}$$

The function $I(x)$ can be derived purely empirically from the function $E(x)$ without introducing any additional hypothesis.

The function $E(x)$ can be approximated roughly by interpolating some of its measured values, and Menzel has proposed the following interpolation formula:

$$E(x) = E_0 e^{-\alpha x}, \tag{2}$$

where $E_0 = E(0)$ and α (the gradient) is a constant which is to be determined from the observations. Experience has shown that, within the limits of measurement accuracy, this simple exponential formula gives a satisfactory description for most of the lines in the flash spectrum. In other cases the observations can be satisfactorily represented by a combination of two exponential terms,

$$E(x) = E_1 e^{-\alpha_1 x} + E_2 e^{-\alpha_2 x}, \tag{3}$$

and it is only seldom that a third term is necessary.

If we insert (2) into (1), then we obtain

$$I(x) = \alpha E_0 e^{-\alpha x} = \alpha E(x), \tag{4}$$

that is, the emission intensity $I(x)$ of the chromosphere at a given height x can be determined by multiplying the total emission of the chromosphere above this level times the intensity gradient α. It should be noted especially that $I(x)$ is in this case expressed by the same exponential function as $E(x)$.

If the chromosphere is transparent to radiation at the wavelength of the given line, then from the intensity I we can compute the number of emitting atoms N_s in a column 1 cm^2 in cross section, parallel to the line of sight, from the expression

$$I = \frac{N_s A_{sr} h\nu}{4\pi}, \tag{5}$$

where A_{sr} is the Einstein coefficient for the spontaneous-transition probability. In the given case the variation in line intensity with the distance from the solar limb corresponds to the variation in the number of atoms N_s. If the line in question is characterized by appreciable self-absorption, then the measured intensities must be reduced to an optically thin layer (see §24).

Let us assume that S is the surface of the sun, with the solar center at C (Figure 33), T is the line of sight, M is the point of contact of the line of sight T with the lunar limb at the given moment, and P is an "instantaneous" point on the line of sight. In addition, we denote the solar radius as R, the heights of points M and P above the solar surface as x and h, and the distances from the center as $r = R + x$ and $\varrho = R + h$.

If $n(\varrho)$ is the number of atoms per cm^3 at a distance ϱ from the center of the sun, while $N(r)$ is the number of atoms in a column 1 cm^2 in cross

section parallel to the line of sight at a distance r from the center of the sun, then according to Figure 33 these two quantities are related by the following expression:

$$N(r)=2\int_{r}^{\infty}\frac{n(\varrho)\,\varrho\,d\varrho}{\sqrt{\varrho^2-r^2}}. \tag{6}$$

If we now insert $\varrho = R + h$ and $r = R + x$, then since h and x for the chromosphere are negligible in comparison with R we obtain

$$N(x)=\sqrt{2R}\int_{x}^{\infty}\frac{n(h)\,dh}{\sqrt{h-x}}. \tag{7}$$

This is an Abel integral equation, the general solution of which has the form

$$n(h)=-\frac{1}{\pi\sqrt{2R}}\frac{d}{dh}\int_{h}^{\infty}\frac{N(x)\,dx}{\sqrt{x-h}}. \tag{8}$$

Let us next assume that the number of atoms N decreases exponentially with the distance from the solar limb:

$$N(x)=N^0e^{-\alpha x}, \tag{9}$$

where N^0 is the value of N at $x = 0$. Then, when (9) is substituted into (8), we find that the concentration of atoms is

$$n(h)=n^0e^{-\alpha h}, \tag{10}$$

where n^0 is the atom concentration at the base of the chromosphere, defined as

$$n^0=\sqrt{\frac{\alpha}{2\pi R}}N_0. \tag{11}$$

Conversely, by substituting (10) into (7) we can find formula (9) for $N(x)$. Therefore, if the distribution of atoms over the height follows the exponential law (10), then the number of atoms in the line of sight varies with height according to an analogous formula with the same gradient, and vice versa. It is evident that this conclusion will also apply for the two-term (or three-term) exponential formulas.

Using formulas (5) and (4), we can express N^0 directly in terms of E_0, that is, in terms of the total emission of the chromosphere:

$$N^0=\frac{4\pi\alpha E_0}{A_{sr}h\nu}, \tag{12}$$

so that, taking into account (11), we obtain

$$n^0 = \frac{4\pi\alpha^{3/2}E_0}{A_{sr}h\nu\sqrt{2\pi R}}. \qquad (13)$$

Finally, if we integrate (10) from $h = 0$ to $h = \infty$, we will find the number of atoms over 1 cm^2 of the base of the chromosphere:

$$N' = \frac{n^0}{\alpha}. \qquad (14)$$

Table 3 gives the values of $\lg E_0$ ($erg \cdot sec^{-1}$) and the gradients α (cm^{-1}) for several chromospheric lines. The values were averaged by Vyazanitsyn from observation results for the solar eclipses of 1941, 1945, and 1952 /6/ (see Figure 34).

TABLE 3

Line	$\lg E_0$	$\alpha \cdot 10^8$	Line	$\lg E_0$	$\alpha \cdot 10^8$
H_α 6562.8	15.69	0.82	Ba+ 4554.1	13.45	1.79
H_β 4861.5	15.48	1.10	Sc+ 4246.9	13.44	1.62
H_γ 4340.6	15.16	1.17	Mn 4034.5	13.31	1.44
H_δ 4101.8	14.97	1.23	Mn 4033.1		
H_8 3889.2	14.73	1.18	Mn 4030.8		
H_9 3835.5	14.67	1.41	Cr 4254.4	12.86	1.60
H_{10} 3798.0	14.16	1.30	Fe 4383.5	13.19	1.77
H_{11} 3770.7	14.01	1.32	Fe 4063.6	13.27	2.10
H_{12} 3750.2	13.90	1.40	Fe 4045.8	13.57	2.10
H_{13} 3734.4	14.06	1.53	Fe 3859.9	13.59	1.55
H_{14} 3722.0	13.86	1.46	Fe+ 5018.4	13.62	1.57
H_{15} 3712.1	13.64	1.49	Fe+ 4924.0	13.43	1.66
H_{16} 3703.9	13.53	1.47	Fe+ 4583.9	13.32	1.92
H_{17} 3697.2	13.13	1.35	Fe+ 4522.7	13.12	1.39
H_{18} 3691.6	12.99	1.51	Fe+ 4233.2	13.19	1.61
He 5875.6	15.06	0.86	Ti+ 4572.0	13.13	1.99
He 4713.2	13.30	0.88	Ti+ 4563.8	13.51	3.30
He 4471.5	13.85	0.92	Ti+ 4549.6	13.31	1.65
He 4026.3	13.47	0.83	Ti+ 4534.0	13.39	2.33
He+ 4685.8	12.15	0.44	Ti+ 4501.3	13.09	1.63
Ca 4226.7	13.41	1.70	Ti+ 4443.8	13.32	2.11
Ca+ 3968 7	15.39	0.82	Ti+ 4395.1	13.51	2.07
Ca+ 3933.9	15.48	0.79	Ti+ 4300.0	13.28	1.91
Sr+ 4215.7	13.80	1.46	Ti+ 4294.1	13.87	2.58
Sr+ 4077.8	14.19	1.49	Ti+ 4290.2		
Mg 5183.6	14.05	1.62	Ti+ 3913.6	13.35	1.35
Mg 5172.6	14.22	1.68	Ti+ 3900.5	13.43	1.51
Mg 5167.4			Ti+ 3761.9	14.04	1.11
Na 5896.0	14.32	2.36	Ti+ 3761.3		
Na 5889.1			Ti+ 3759.3		
Ba+ 4934.1	13.37	2.26	Ti+ 3685.2	13.88	1.20

Several interesting features of the chromospheric spectrum, some of which were discussed above in §20, emerge from Table 3:

1) helium lines are observed, and the hydrogen lines are more numerous than in the solar [photospheric] spectrum;

2) the H and K lines of ionized calcium, and also the lines of helium (particularly He^+ 4686), have very small gradients;

3) the gradients increase with the number of the line in the Balmer series;

4) the gradient of a neutral element is invariably higher than the gradient of the corresponding ionized element, which indicates that the ionization intensifies with height;

5) all the observed emission gradients are considerably lower than those expected for the gradients of density decrease in the presence of gravitational equilibrium, even the hydrogen density gradient ($\alpha = 6.6 \cdot 10^{-8}$);

6) there is no correlation between the gradient and the atomic weight.

The study of the height distribution of atoms in various excitation states is of interest from the point of view of the variation in the excitation factors over the height in the chromosphere, provided the chromosphere is regarded as a homogeneous layer. However, there is another subject which is of considerable independent interest, namely the distribution of all the atoms (or ions) of a given element over the height, or, which is practically the same thing, the distribution of the atoms in the ground state. The excitation of emission in lines with low excitation potentials (at least up to 5 or 7 eV) can be definitely attributed to solar radiation. In this case, it is easy to derive the following formula for computing the number of atoms N_1^0 in the ground state according to the number of excited atoms (at the base of the chromosphere):

$$N_1^0 = N_s^0 \left(\frac{2he^2}{\lambda^5} \frac{g_1}{g_2} \right) \frac{1}{I_\lambda} = \frac{8\pi e}{\lambda^4} \frac{g_1}{g_2} \frac{\alpha E_0}{A_{s1}} \frac{1}{I_\lambda}, \qquad (15)$$

where I_λ is the intensity of the exciting radiation, which is almost constant throughout the chromosphere.

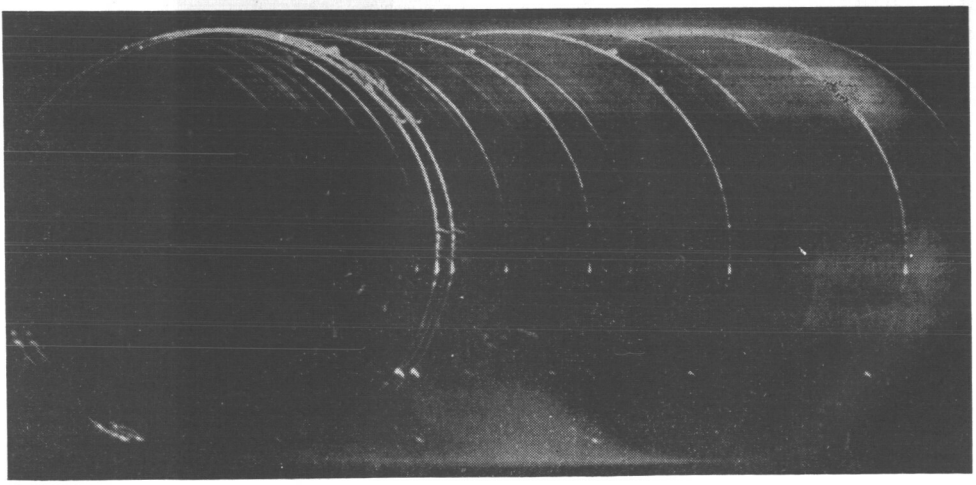


FIGURE 34

Table 3 contains the resonance lines of Na, Ca, Sr^+, and Ba^+, for which the transition-probability coefficients are known. Thus, we can use the observed gradients and intensities to compute from formula (15) the values of N_1^0 for these elements. The results of these computations are given in Table 4, which also lists the values of A_{sr} and I_λ. Here, the intensity of the exciting radiation was assumed to equal the central intensity of the corresponding Fraunhofer line, taken from the Utrecht photometric atlas of the solar spectrum /19/ and expressed in absolute units, a dilution factor of 1/2

being introduced. The T_{ex} column in Table 4 gives the corresponding excitation temperatures. It should be noted that, if in the lower layers of the chromosphere the lines are widened considerably by turbulence, then a somewhat higher I_λ should be taken and, accordingly, the computed numbers of atoms should be reduced. Successive ionization potentials of the elements being considered are listed in Table 5.

TABLE 4

Line	$A_{sr} \cdot 10^{-7}$	$I_\lambda \cdot 10^{-13}$	$\lg N_1^0$	T_{ex}	$\lg N^0$	$\lg n^0$	$\lg N'$
Ca 4227	2.83	4.70	14.41	3460°	18.52	8.81	16.59
Sr+ 4216	4.95	21.2					
Sr+ 4078	3.25	18.7	14.52	4085°	14.55	4.82	12.65
Ba+ 4934	9.15	25.0					
Ba+ 4554	10.71	40.4	13.23	4130°	13.44	3.78	11.47
Na 5896	6.33	48.4					
Na 5889	6.45	45.1	13.34	3710°	18.29	8.66	16.28

TABLE 5

Element	χ_0	χ_1	χ_2
Na	5.12	47.0	
Ca	6.09	11.82	5.0
Sr	5.67	10.98	
Ba	5.19	9.96	

According to Saha's formula, almost all the Na atoms are singly ionized at all the temperatures reached in the chromosphere. The neutral states of the other three elements can be neglected with equal certainty, but this does not apply to the doubly ionized state. However, since the H and K lines of Ca^+ in the chromospheric spectrum are exceptionally bright, it is highly probable that most of the calcium atoms are singly ionized. A comparison of the number of Ca^+ ions at the base of the chromosphere ($\lg N_1^0 = 18.52$) found by us (see §24, Table 9) with the number of Ca atoms ($\lg N_1^0 = 14.41$) makes it possible to calculate the ionization temperature for calcium. If we assume the electron concentration given by Cillie and Menzel ($\lg n_e^0 = 11.59$), then we find that $T_i = 5240°$. Then, for this temperature we can use Saha's formula and the data in Table 4 (column 4) to calculate the number of Na^+ ions, as well as the number of doubly ionized atoms of calcium, strontium, and barium. When these values are added to the original numbers listed in Table 4, we will have the amount of all the atoms of Na, Ca, Sr, and Ba in a column 1 cm^2 in cross section (parallel to the line of sight) at the base of the chromosphere. These quantities ($\lg N^0$) are also listed in Table 4. The corresponding atom concentrations n^0 at the base of the chromosphere and the number of atoms in 1 cm^2 of the chromosphere base, calculated from (13) and (14), are listed in the final two columns.

§24. Self-absorption in the chromosphere. For most chromospheric lines the self-absorption is negligible, at least to a first approximation. However, this effect can have a substantial influence for some of the brighter lines and for the deeper levels of the chromosphere. For example, this is undoubtedly the case for the H and K lines of Ca II, as indicated by

the presence of intense line wings and by reversal in the central parts of these lines in the lower layers. Self-absorption should also be considerable in the first Balmer hydrogen lines, especially in the H_α line. Thus, taking self-absorption into account in these cases is a necessary condition for passing from the measured line intensities to the numbers of atoms; it is, however, a very laborious task and it complicates the interpretation of spectrophotometric measurements considerably. Let us now look at this problem as it applies to the interpretation of eclipse observations of the chromosphere, which yield absolute line intensities.

We assume that I_ν is the emission of the chromosphere at a frequency ν at some point having an optical thickness τ_ν parallel to the line of sight, t_ν is the optical thickness of the chromosphere in the same direction, $\varkappa_\nu$ is the absorption coefficient, j_ν is the emission coefficient, and $J_\nu = \frac{j_\nu}{\varkappa_\nu}$ is the so-called yield function. If we solve the equation of radioactive transfer

$$\frac{dI_\nu}{d\tau_\nu} = -I_\nu + J, \tag{16}$$

then we obtain

$$I_\nu = \int_0^{t_\nu} J_\nu e^{-\tau_\nu} d\tau_\nu, \tag{17}$$

so that we can integrate over the frequency to find the expression for the intensity of a chromospheric line in the presence of self-absorption:

$$I = \int d\nu \int_0^{\tau_\nu} J_\nu e^{-\tau_\nu} d\tau_\nu. \tag{18}$$

The main difficulty involved in allowing for the self-absorption is that the yield function J_ν is unknown, chiefly as a result of the sparseness of available data on the mechanism of atomic excitation in the chromosphere. The absorption coefficient is generally assumed to vary with frequency mainly due to natural damping and due to the Doppler effect (thermal velocity and turbulence). This coefficient also depends on the concentration of absorbing atoms. All these quantities differ for different levels in the chromosphere, but our information on the way in which they vary over the height (for example, the turbulent velocities at various heights) is still highly inaccurate. This indicates quite well how difficult it is to take self-absorption strictly into account. The physical difficulties, moreover, are augmented by purely mathematical complications. For example, the formula for the absorption coefficient when the two previously mentioned line-broadening factors are present has a complicated integral form. Also, considerable difficulty is involved in calculating the integral in the first part of (18) for the general case.

In order to simplify the problem, let us assume that coefficients j_ν and $\varkappa_\nu$, and thus also the yield J_ν, remain constant along the line of sight. Then, integrating (18) over τ_ν and substituting $t_\nu = \alpha_\nu N_r$, where α_ν is the atomic absorption coefficient and N_r is the number of absorbing atoms in the line

of sight, we obtain

$$I = \int J_\nu (1 - e^{-\alpha_\nu N_r})\, a\nu. \tag{19}$$

Regardless of the mechanism of atomic excitation (the scattering of solar radiation, fluorescence, electron-impact excitation), the function J_ν will vary relatively little with frequency inside the line, so that we may set $J_\nu = \bar{J}$, where $\bar{J}$ is the average value of J_ν. Finally, we have

$$I = \bar{J} \int (1 - e^{-\alpha_\nu N_r})\, d\nu. \tag{20}$$

Theoretical studies of the chromospheric self-absorption, as well as ways of taking it into consideration, can be found in the works of Pannekoek /20/, Menzel /1/, and Cillié and Menzel /4/. It was assumed that excitation of chromospheric atoms is due to the absorption of photospheric radiation and that in a given line $J_\nu(x)$ = constant = $\bar{J}$ for all heights in the chromosphere. Menzel applied the general expression for the absorption coefficient (natural damping plus Doppler effect).

By analogy with the theory of growth curves for absorption lines (see Chapter II), we may use equation (20) to construct the corresponding curves for emission lines, which give the variation of the equivalent width

$$W = \frac{I}{\bar{J}} \tag{21}$$

as the number of absorbing atoms N_r in the line of sight is increased. Here, in each case a definite mean velocity of the atoms is assumed, or, equivalently, a definite ratio of the natural and Doppler half-widths $a = \frac{\Delta\lambda_N}{\Delta\lambda_D}$. In order to calculate $I/\bar{J}$ Menzel proposed four alternative forms of formula (21), corresponding to different optical thicknesses $a_1 N_r$ /4/.

When the densities increase exponentially with the height into the chromosphere, the number of atoms in the line of sight also decreases as an exponential function of the distance from the solar limb (see §23). On the basis of this fact Menzel represented the curve of growth as a function of the equivalent width for a definite gradient a but for an arbitrary line-of-sight number of atoms N_r^0 near the solar limb. The quadrature integration of this curve for various heights x gives the growth curve, which relates the equivalent widths of the chromospheric emission above a certain level to the height of this level.

By means of successive approximations (a variation of the parameter a), we find the curve which best fits the observed $E(x)$ curve. Since a variation of N_r^0 produces just a lateral translation of the curve, while a variation of $\bar{J}$ displaces it vertically, the shape of the curve remaining unchanged in either case, this method enables a determination of the three independent parameters a, N_r^0, and $\bar{J}$.

Taking into account the self-absorption for hydrogen lines. The emission gradients of the first Balmer lines decrease in the lower layers of the chromosphere, and this phenomenon intensifies as we pass from H_δ to H_α (Figure 35). This should apparently be attributed to self-absorption. Without discussing the significance of a certain reduction of the gradients of these lines in the upper layers, let us apply the preceding

method for taking the self-absorption into account in order to determine the following physical quantities in the chromosphere: the number of hydrogen atoms N_2^0 at the second level in a column 1 cm^2 in cross section along the line of sight at the base of the chromosphere, the density gradient α, and the "excitation function" $\bar{J}$. Let us set $\bar{J} = P\Phi$, where P is Planck's function and $\Phi < 1$ is a coefficient allowing for the variation of emission with depth (assumed constant for simplicity).

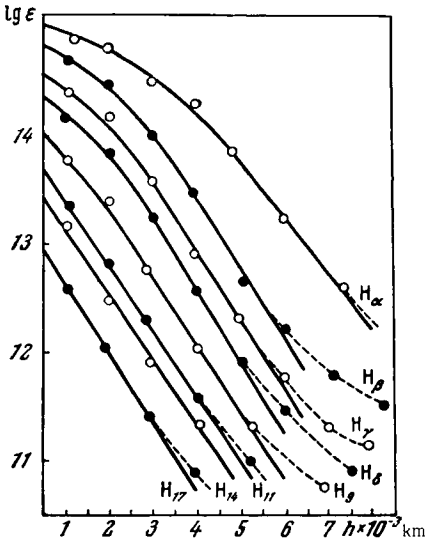


FIGURE 35

The damping constant in this case (which is virtually constant for all the Balmer lines) is $6.6 \cdot 10^8$ sec^{-1}. If we take v = 9 km/sec as the average velocity of the hydrogen atoms, which corresponds to T = 4830°, then we obtain for the lines from H_α to H_δ the following range of the parameter: $a = \frac{\Delta\lambda_N}{\Delta\lambda_D} = 0.008$ to 0.005. Next, using Held's tables /21/, we can find the equivalent widths W for several values of N_2. Then, after selecting a value of α, successive approximations are used to find for every line which distribution curve of the intensities (equivalent widths) $\int W\,dx$ over the height best fits the observational results. The theoretical curves for the H_α, H_β, H_γ, and H_δ lines are shown in Figure 35, and the corresponding density gradients are listed in Table 6. A comparison of the readings along each of the coordinate axes for the observed and theoretical curves (when these agree, as in Figure 35) gives the constants N_2^0 and $\bar{J}$ (in erg·cm^{-2}·sec^{-1} per Å of the spectrum per unit solid angle), which are also listed in Table 6.

As expected, the emission gradients of the first Balmer lines, reduced to an optically thin layer, are higher than the observed values (see Table 3), and they approach the gradients of the upper lines in the series, which are free of self-absorption. The average gradients of the lines from H_α to H_δ practically coincide with the average gradients of the lines from H_{10}

to H_{18} in Table 3 ($\bar{\alpha} = 1.43 \cdot 10^{-8}$). The systematic variation of $\lg N_2^0$ from H_α to H_δ should be attributed to the extreme schematization of the chromospheric conditions which is typical of this means of taking into account the self-absorption. The function $\bar{J}$ does not vary much from H_α to H_δ. The corresponding values of the temperature (T) calculated from Planck's formula (Table 6) establish the minimum temperature of the excitation source ($\Phi = 1$). Since in the chromosphere $\Phi < 1$, we conclude that the hydrogen in the chromosphere is excited by some source which is more powerful than just the photospheric radiation.

TABLE 6

Line	v=9 km/see				v=9 km/sec, v'=15 km/sec		
	$\alpha \cdot 10^8$	$\lg N_2^0$	$\lg \bar{J}$	T	$\lg N_2^0$	$\lg \bar{J}$	T
H_α	1.30	14.53	6.54	6450°	14.82	6.25	5450°
H_β	1.45	14.80	6.74	6700	15.09	6.45	5850
H_γ	1.45	14.84	6 69	6550	15.13	6.40	5800
H_δ	1.60	15.40	6.38	5750	15.69	6.09	5200
Average	1.45	14.89	—	6350°	15.18	—	5550°

The existence of chromospheric turbulence was disregarded in the previous computations. On the basis of an estimate of the maximum deviation (from the above results) of $\lg N_2^0$ and $\lg \bar{J}$ when turbulence is taken into account, we obtain $v' = 15$ km/sec. Thus, the resulting velocity of the atoms is 17.5 km/sec. Since the shape of the theoretical curve is determined mainly by the gradient, we can retain the previous values of the gradients for all four lines. The final results of the calculations for this situation are given in the second half of Table 6.

Taking into account the self-absorption for the H and K calcium lines. Let us now make an analogous calculation for the H and K lines of Ca^+. The damping constant for these lines is $1.55 \cdot 10^8$ sec^{-1}, and the thermal velocity of calcium atoms at $T = 4830°$ is 1.4 km/sec. Consequently, $a = \Delta\lambda_N / \Delta\lambda_D \approx 0.007$. Next, using successive approximations we obtain the curve of growth. The corresponding values of the gradient α, the number of atoms N^0 in the ground state, and the excitation function $\bar{J} = P\Phi$, as well as the Planck temperatures T derived independently accordingly to the K and H lines, are given in the left half of Table 7.

TABLE 7

Line	v=1.4 km/sec				v=1.4 km/sec, v'=7 km/sec			
	$\alpha \cdot 10^8$	$\lg N^0$	$\lg \bar{J}$	T	$\alpha \cdot 10^8$	$\lg N^0$	$\lg J$	T
K Ca^+	2.1	18.50	6.20	5450°	2.5	20.20	5.50	4400°
H Ca^+	2.1	18.53	6.16	5350	—	—	—	—
Average	2.1	18.52	—	5400°	—	—	—	—

If we assume a turbulent velocity of ~ 7 km/sec, then we find (using successive approximations) the curve of growth for Ca^+K whose constants are listed in the right half of Table 7. The low excitation temperature is consistent with the universally accepted assumption that the chromospheric emission in the Ca^+H and K lines is due to resonance scattering of the photospheric radiation.

§25. The temperature and the electron concentration. Chromospheric models. The physical state of the chromosphere, which is not in thermodynamic equilibrium, is described in terms of various arbitrary "temperatures." These temperatures, which are actually parameters specifying various properties of a system in thermodynamic equilibrium, are: the electron temperature T_e, which generally equals the kinetic temperature T_k, the excitation temperature T_{ex}, the ionization temperature T_i, etc.

The electron temperature can be found from the distribution of energy in the Balmer continuum. This emission, which arises due to proton-electron recombination /22/, is equal to

$$2.63 \cdot 10^{-33} n_p n_e T_e^{-3/2} e^{(\chi_2 - h\nu)/kT_e} d\nu \text{ erg} \cdot \text{cm}^{-3} \cdot \text{sec}^{-1}, \qquad (22)$$

where n_p and n_e are the concentrations of protons and electrons, respectively, and $\chi_2 = 3.40$ eV (second-level binding energy for hydrogen).

After measuring from the 1932 eclipse photographs the emission at $\lambda\lambda$ 3640 and 3500 Å for two chromospheric layers (at heights of 670 and 900 km, respectively), Cillié and Menzel obtained values of $T_e = 5250°$ and 4020° from formula (22).* Previously, Davidson, Minnaert, Ornstein, and Stratton (for the 1928 eclipse) found $T_e = 3600$ to 4000° (at heights from 400 to 700 km), while Zanstra (1950) calculated $T_e = 6000°$ from the measurements of these authors. V. A. Krat (for the 1945 eclipse) found $T_e = 6200°$ for the upper chromosphere (7000 to 8000 km).

Athay, Billings, Evans, and Roberts /9/, using photographs taken during the 1952 eclipse, measured the intensities of the Balmer continuum at four wavelengths and at different heights. From formula (22) they found the corresponding values of T_e, which increases from 4300° at 400 km to 11,100° at 2400 km, the average for this layer being $T_e = 6200°$. If the chromosphere is assumed to be isothermal, which, as pointed out by these authors, is definitely inconsistent with their measurements, then the average temperature over the given layer is found to be 8000°.

The kinetic temperature can be determined from line profiles which show Doppler broadening. In the absence of self-absorption, the Doppler profile of an emission line is given by the formula

$$I(x) = \frac{cI_0}{\lambda V \sqrt{\pi}} e^{-\frac{c^2(\lambda - \lambda_0)^2}{\lambda_0^2 V^2}}, \qquad (23)$$

where

$$V^2 = v^2 + \xi^2 \qquad (24)$$

* Actually, temperatures of $\sim 10,000°$ were obtained, but subsequently it was established that an error had been made in the calculations /3/.

$v = \sqrt{\frac{2kT_K}{m}}$ being the most probable velocity of thermal motion of the atoms and ξ being the turbulent velocity. By measuring the line profiles, we can find V and thus also T_K and ξ. Obviously, if $\xi \gg v$, then for equal intensities the lines of all elements must have almost identical profiles, whereas if $v \gg \xi$ the line width must decrease with an increase in atomic number.

Redman determined the chromospheric temperature from the weak sharp lines of hydrogen, helium, and the rare earths, using a flash spectrum from the 1940 eclipse /23/. He found that the profiles of these lines correspond to a Gaussian distribution of the velocities of the emitting atoms, where the mean velocities for the different elements have the following values:

H, 13.5; He, 8.0; La, 1.6; heavy elements, 1.75 km/sec.

The velocities were thus approximately inversely proportional to $\sqrt{m}$, which would indicate mostly thermal velocities. More precisely, it was established that in the lower chromosphere (1500 km) the turbulent velocity does not exceed ~ 1 km/sec and the temperature T_K = 35,000°. However, from measurements during the 1952 eclipse Redman found turbulent velocities up to 2 km/sec and T_K values of at most 10,000°, with possible values of 6000° /24/.

Similarly contradictory results were obtained for the excitation and ionization temperatures of elements in the chromosphere. Metals are excited and ionized at $\sim$4000 to 6000°, and the excitation temperature of orthohelium is about the same. On the other hand, to explain the observed hydrogen emission (by applying the combined Boltzmann-Saha formula) a temperature of $\sim$10,000° is required, while ionization equilibrium of helium corresponds to $\sim$20,000°. From an analysis of the Balmer decrement, Thomas (1950) found T = 35,000°. On the other hand, Woolley has shown that for this chromospheric temperature the ionization in the earth's atmosphere would have to be much higher than that actually observed.

The fact that T_e, T_K, T_{ex}, and T_i in the chromosphere are all different shows that the chromosphere is far from being in thermodynamic equilibrium. However, there is still no comprehensive theory of the physical state of the chromosphere, and an interpretation of the various observations from a single viewpoint is a difficult problem, particularly since the observational data themselves are known to be contradictory. Whereas some observations (line widths, density gradients, helium ionization, and Balmer decrement) indicate a high chromospheric temperature, other observations (energy distribution in the Balmer continuum, excitation of helium and of metals, radio observations, and ionospheric disturbances) indicate a low temperature.

Another important conclusion concerning the temperature regime of the chromosphere, one reached only during the last 10 or 15 years, points toward the existence of a positive temperature gradient in the chromosphere. This is consistent, of course, with the high temperature of the solar corona (see Chapter VII), and also with the existence of a maximum for helium lines in the chromosphere (see §21).

Another important characteristic of the state of the chromosphere, in addition to the temperature, is the electron concentration. Various techniques have been proposed for its determination. Measurements of the

intensity of the Balmer continuum near the limit of the series (λ = 3640 Å) at various heights in the lower chromosphere were made by Cillié and Menzel. These authors then used formula (22) to find

$$n_p n_e T_e^{-3/2} = 1.45 \cdot 10^7 \cdot e^{-1.54 \cdot 10^{-8}}. \quad (25)$$

Then, by setting $n_p = n_e$ and T_e = 10,000°, they obtained for the base of the chromosphere $\lg n_e$ = 11.59 (or an electron pressure of 0.55 dyne/cm^2). It is important to note that the concentration n_e is not very sensitive to changes in T_e; for example, for T_e = 5000° we obtain $\lg n_e$ = 11.36. The electron-concentration gradient according to Cillié and Menzel is $0.77 \cdot 10^{-8}$ cm^{-1}

For this same purpose Wildt /3/ used the formula of Inglis and Teller /25/:

$$\lg c = 23.26 - 7.5 \lg n, \quad (26)$$

which gives the principal quantum number n of the last resolvable line in the series as a function of the concentration of charged particles c (the lines blend due to the broadening of the statistical Stark effect). For temperatures less than $10^5/n$, the quantity c is the number of ions plus the number of electrons ($c = 2n_e$), while for temperatures over $10^5/n$ the electrons do not contribute to the broadening ($c = n_e$). According to the Mitchell catalog the last detectable Balmer line in the flash spectrum is n = 37 and the last detectable line in the Paschen series is n = 40; consequently, Wildt found $\lg n_e$ = 11.50 and 11.24, respectively (for the 500 km height), which is in close agreement with the result of Cillié and Menzel. Naturally, the insufficient resolving power of the spectrograph and the Doppler broadening of the lines will increase the determined value of n_e somewhat. The electron-concentration gradient according to Wildt is $0.46 \cdot 10^{-8}$ cm^{-1}. Previously (1938), Pannekoek had used an analogous method to obtain the value $\lg n_e$ = 11.78.

According to the previously mentioned measurements of American astronomers during the 1952 eclipse /9/, $\lg n_e$ varies from 11.62 at 400 km down to 10.97 at 2400 km; the corresponding electron-concentration gradient is $\sim 1.05 \cdot 10^{-8}$cm^{-1}.

Finally, let us consider the electron concentrations at various heights in the chromosphere established by Vyazanitsyn from intensity measurements of the Balmer lines from H_α to H_{17} during the solar eclipses of 1941, 1945, and 1952 /6/. Vyazanitsyn used the combined Boltzmann-Saha formula for hydrogen:

$$\lg n_p n_e = \lg \frac{n_s}{g_s} + \frac{3}{2} \lg T_e - \frac{5040}{T_e}(\chi_0 - \chi_s) + 15.684, \quad (27)$$

where n_p is the proton concentration, and χ_0 is the ionization potential of hydrogen. The initial data (the average values for the three eclipses) were the numbers of hydrogen atoms in the 2nd state at various heights (as determined from the line intensities) and the corresponding Boltzmann temperatures averaged over all the lines. These quantities are listed in the first two rows of Table 8. The third row gives the atom concentrations at the various heights calculated from formula (13). According to Boltzmann's formula, as we approach the series limit ($\chi_s - \chi_2 \to 3.38$ eV) with T

constant, the number of atoms in the s state $\frac{n_s}{g_s}$ per quantum cell approaches a certain limit. If we insert this limit value of $\frac{n_s}{g_s}$ into (27) and set $\chi_s = \chi_0$, then we find the value of $\lg n_p n_e$. If it is taken into account that the electrons are mainly products of the ionization of hydrogen (for which $n_p = n_e$), then we can find the electron concentration n_e. The resulting values of $\frac{n_s}{g_s}$ and n_e are given in Table 8.

TABLE 8

h, km	500	1500	2500	3500	4500	5500	6500	7250
$\lg N_2$	14.58	13.94	13.32	12.68	12.06	11.42	10.79	10.32
T_e	6170°	6240°	6260°	6280°	6160°	6330°	6710°	5660°
$\lg n_2$	4.84	4.20	3.58	2.94	2.32	1.68	1.05	0.58
$\lg (n_s/g_s)$	1.18	0.56	1.96	1.33	2.65	2.08	3.61	4.66
$\lg n_e$	11.28	10.97	10.67	10.36	10.01	9.73	9.52	8.99
$\lg n_e^*$	11.22	10.92	10.62	10.31	9.95	9.67	9.45	8.90

Figure 36 shows $\lg n_e$ as a function of the height h, and the graph indicates that n_e decreases exponentially with height. The solution of the conditional equations specifying the position of the straight line on the graph, using the method of least squares, gives the following values of the electron concentration at the base of the chromosphere and the electron-concentration gradient: $\lg n_e^0 = 11.45$ and $\alpha = 0.724 \cdot 10^{-8}$ cm^{-1}. Consequently, the value of $\lg n_e^0$ is in excellent agreement with the results of other authors who used entirely different methods. The electron-concentration gradient, moreover, agrees with the result of Cillié and Menzel ($0.77 \cdot 10^{-8}$) for the lower chromosphere.

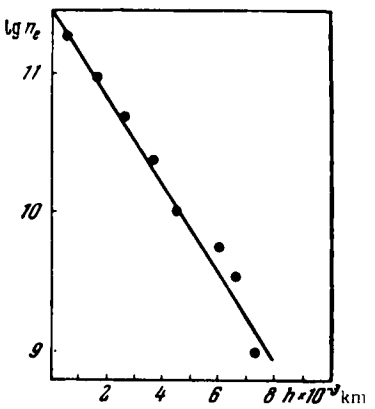


FIGURE 36

The last row of Table 8 gives the electron concentrations $\lg n_e^*$ for $V = 175$ km/sec, computed in a similar way. The following values were obtained in this case for the electron concentration at the base of the chromosphere and the electron-concentration gradient: $\lg n_e^0 = 11.41$ and $\alpha = 0.734 \cdot 10^{-8}$ cm^{-1}. Consequently, allowing for turbulence ($\xi = 15$ km/sec) practically did not affect the values of the electron concentration and the gradient of its decrease with height in the chromosphere.

Many attempts have been made recently to construct a chromospheric model which would explain the observational data. The main feature of these models which distinguishes them from earlier ones is the recognition that the chromosphere is nonuniform, namely that cold and hot regions are present in it. No theory representing the chromosphere as a homogeneous gaseous layer with a definite height distribution of n_e and T_e could explain

the simultaneous existence in the flash spectrum of the lines of metals, hydrogen, helium, and ionized helium, whose excitation potentials differ sharply from one another. The nonuniformity of the chromosphere is verified directly by observations of the spicules and the chromospheric fine structure, which show up on spectroheliograms in the light of the H_α and H and K Ca II lines when the atmospheric conditions are favorable. Models of a nonuniform chromosphere have been suggested by Giovanelli /26/, Hagen /27/, Woltjer /28/, Athay and Menzel /29/, Shklovskii and Kononovich /30/, and V. A. Krat and his co-workers /31, 32/. The last three models deserve a more detailed consideration.

Athay and Menzel, after processing observational data for the chromospheric helium spectrum of the 1952 total solar eclipse, obtained the following expression for the intensity per unit volume as a function of the height of the chromosphere:

$$\mathcal{E}_n = \frac{1}{K_n}(2.7\cdot 10^{-2}e^{-1.26h} - 4.3e^{-3.7h} + 15e^{-4.6h} - 6.8\cdot 10^{2}e^{-9.2h}), \qquad (28)$$

where $\frac{1}{K_n}$ is a numerical factor which is different for different lines. The equation is valid for $h > 1000$ km. From equation (28) it is possible to determine the populations of various levels, since self-absorption is virtually nonexistent.

Let us write the Boltzmann-Saha equation in the form

$$n_n = 4.14\cdot 10^{-16} n_{II} n_e T_e^{-3/2} \frac{g_n}{4}\cdot 10^{5040\frac{\chi_n}{T}}, \qquad (29)$$

where n_{II} is the concentration of helium ions, g_n is the statistical weight, and χ_n is the excitation potential of level n. If we write this equation for the various levels, inserting in each case the level population n_n found from the observations, then we obtain a graph of n_n as a function of χ_n, from which the following expression may be derived:

$$n_{II} n_e = 1.3\cdot 10^{14}, \qquad (30)$$

which gives the value of the right side of equation (29) for $\chi_n = 0$.

The authors then suggest that cold and hot regions exist in the chromosphere, with fractional volumes a_1 and $a_2 = 1 - a_1$, respectively, so defined that the thermal emission arises from regions of type 1. Equation (29) can now be rewritten as

$$n_{II} n_e T_1^{3/2} = \frac{1.3\cdot 10^{14}}{a_1}\frac{\Phi(h')}{\Phi(1.5)}, \qquad (31)$$

where $\Phi(h')$ is defined in equation (28). Equation (31) then may be combined with Boltzmann's equation to give n_e and T_1. Analogously, we can determine the characteristics for the regions of the other type, provided we introduce observational data for the continuum in the regions of λ 3640 and λ 4700 and the intensities of the He II line at λ 4686.

The result is a height distribution of the electron temperatures for the two regions (hot and cold), as shown in Figure 37. The cold regions are identified with spicules and the hot regions with the interspicular matter. Figure 37 shows that the boundary of the lower chromosphere is located

at 3500 km, where the cold regions correspond to T_e = 6300° and the hot regions to $T_e \approx 19{,}000°$. Above 3500 km, the temperature of the cold regions increases to 19,000°, and that of the hot regions to 100,000°. Figure 38 shows a schematic model of the different chromospheric regions.

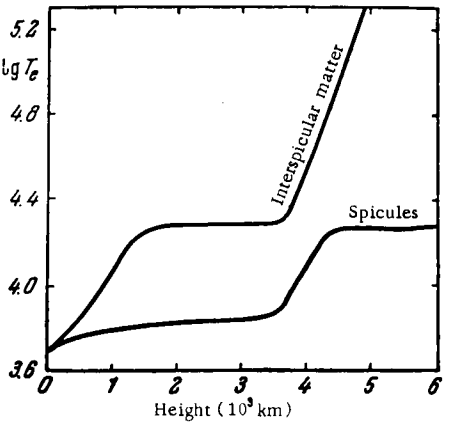


FIGURE 37

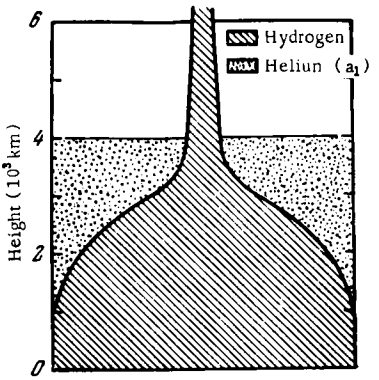


FIGURE 38

However, as shown by Shklovskii and Kononovich /30/, the model of Athay and Menzel is inconsistent, especially with radio observations. They compared the data of radio observations in the centimeter range, obtained during the same year (1952), with the model of Athay and Menzel, and found that these data give an equivalent temperature of the chromosphere of 11,250° for λ = 3.2 cm and 38,600° for λ = 10 cm. Shklovskii and Kononovich computed the optical thickness of the chromosphere for these wavelengths according to the model of Athay and Menzel for the hot and the cold components, and they found that the spicules should be opaque to radio waves in the centimeter range.

In order to ensure that the radio emission originates in the hot component of the upper chromosphere, on the other hand, the electron concentration must be one third of that prescribed by the model of Athay and Menzel. Moreover, if according to Athay and Menzel T_e = 100,000° in the interspicular matter of the upper chromosphere, then the brightness temperature for λ = 3.2 cm should be 48,000°, whereas the observations give 11,000°. Consequently, the model of Athay and Menzel cannot explain the chromospheric radio emission. In addition, Shklovskii and Kononovich made a rough calculation of the number of quanta emitted by the helium atoms according to the model of Athay and Menzel, and they concluded that the observed order of magnitude of the line intensities for helium cannot be explained by this model either. Finally, the application of Boltzmann's formula to calculate the populations of the various helium levels is not entirely justified.

On the basis of radio data, Shklovskii and Kononovich calculated n_e and T_e for various heights in the chromosphere, and then, on the basis of optical observations, they attempted to construct a chromospheric model which

would explain both the optical and radio observations. These authors also concluded that the chromosphere is made up of hot and cold components, and their model is illustrated in Figure 39. In contrast to the model of Athay and Menzel, the spicules here represent the hot component, which is responsible for the helium emission.

However, both the Athay-Menzel and Shklovskii-Kononovich models are based on calculations of helium ionization and excitation which are too rough, and they frequently use Boltzmann's formula, which is not always applicable. In order to better validate these models, stationarity equations must be constructed for the two components, in which the basic processes leading to a population of the the various levels and to a decay of the excited states are taken into account. Finally, the observed populations must be compared with the theoretical results obtained from the stationarity equations. This procedure is recommended not only for helium but also for hydrogen and the metals.

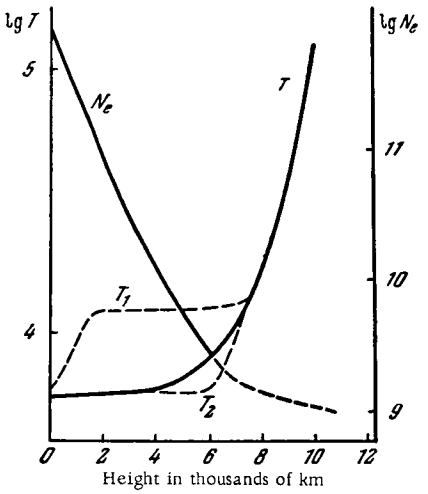


FIGURE 39

V. A. Krat and his co-workers /31, 32/ proposed a different chromospheric model, based on the concept of a dynamic chromosphere consisting of a multitude of filaments with various physical characteristics. Since in the theoretical explanation it is difficult to operate with a continuous spectrum of physical quantities specifying the state of the chromospheric filaments, therefore these quantities were divided arbitrarily into groups characterized by certain average conditions.

The main index from which we can learn about the properties of these filaments is the spectra of the various elements. The filaments in which the lines of ionized helium are the most intense are called subcoronal filaments; for these, T_e = 100,000 to 150,000° and $n_e = 10^{10}$. The so-called helium filaments are cooler, and in them optimum conditions for photohelium emission ($T_e \approx 25{,}000°$ and $n_e = 10^{11}$) are reached. Finally, there are two groups of cold filaments: hydrogen filaments with $T_e = 6000$ to 8000° and $n_H = 10^{13}$, where optimum conditions for hydrogen emission prevail, and metal filaments with $T_e < 6000°$, which are responsible for the emission of the metal lines. Since at the solar limb the different filaments projected onto the observer's visual plane are superimposed, the interpretation of the chromospheric spectrum generally involves certain average quantities, whose distributions can be described by statistical equations. One such equation is the quasi-barometric formula for the distribution of atoms over the height:

$$n = n_0 e^{-\beta h}.$$

When the theory of a homogeneous chromosphere predominated among astrophysicists, this equation was interpreted as a barometric law

describing a density distribution in a gravitational field. Since observations give different gradients β for the lines of different elements, however, such a result was inconsistent with the theory of a homogeneous chromosphere. Consequently, theories of anomalous excitation, anomalous ionization, and anomalous forces maintaining the chromosphere were developed. The new model eliminates all these anomalies, and the quasibarometric formula describes the variation with height not of the atomic concentration but of the concentration of filaments.

The differences between the chromospheric filaments are regarded not as intrinsic properties conserved throughout the entire filament lifetimes but rather as manifestations of various stages of certain elementary dynamic processes. Part of the chromospheric filaments may possibly have intermediate values of β, but values which are not favorable for hydrogen or helium emission and which therefore are difficult to observe. The absence of lines in the spectrum of some particular part of the chromosphere does not indicate a real "cavity" in this region. Nevertheless, together with the various kinds of chromospheric filaments there should also exist chromospheric "cavities," or "windows," as well. The data on radio emission in the centimeter range also indicate this.

The decrease in the intensity of the chromospheric lines with height points not to a change in the atomic concentration but rather to a change in the number of filaments in the line of sight. The fraction of the space within a unit volume element of the chromosphere which is occupied by chromospheric filaments of a certain kind will be called the concentration of these filaments. Let us denote the concentrations of metallic, hydrogen, helium, subcoronal, and intermediate filaments and the concentration of coronal matter as ω_1, ω_2, ω_3, ω_4, ω_5, and ω_6, respectively. Observations show that

$$\omega_i = \omega_{i,0} e^{-\beta_i h}, \tag{32}$$

where $i = 1, 2, 3, 4$, and where

$$\omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5 + \omega_6 = 1. \tag{33}$$

If we take the following observed values for β_i:

$$\beta_1 = \beta_2 = 1.0, \quad \beta_3 = 0.65, \quad \beta_4 = 0.40,$$

and if we measure h not from the level of the photosphere but from the level $h = 1$ (where unity here denotes 1000 km), then we find that $\omega_6 = 1$ for $h = 14$, while for $h = 1$ we have $\omega_5 + \omega_6 = 0$, as well as $\omega_{1,0} = \omega_{2,0} = \omega_{4,0} = 0.2$ and $\omega_{3,0} = 0.4$. Then, by applying equations (32) and (33), we obtain the following limits for the propagation of the various types of filaments:

$$h_1 = h_2 = 5.3, \quad h_3 = 9.3, \quad h_4 = 13.2.$$

In order to verify these estimates, let us now calculate the height at which fluctuations of the hydrogen filaments will result in the jagged edge of the chromosphere apparent in H_α light, that is, the level at which the hydrogen chromosphere starts breaking up into spicules. Although the spicules may exist at all levels of the chromosphere, they mask one another in the lower

layers and create the false impression of a continuous medium. We denote the concentration of hydrogen filaments in a spicule as ω_s, and we assume that one spicule remains in the line of sight at some $h = h_m$. Above h_m separate spicules are observed. If we integrate all the ω_2 along the line of sight, assuming that ω_2 in every chromospheric volume is described by (32), then we find that

$$\omega_s = \frac{\sqrt{2\pi R_\odot}}{\beta_2} \omega_{2,0} e^{-\beta_2(h-1)}. \qquad (34)$$

The quantity $\omega_{2,0}$ cannot differ substantially from ω_s, since for $n = 1$ the chromosphere consists of the same spicules, which occupy the entire space. We may thus take $\omega_{2,0} = \omega_s$. Then, for $\beta_2 = 1.0$ we have

$$66.2e^{-(h-1)} = 1,$$

that is, $h_m = 5.2$. Observations show that this is in fact the height at which the chromosphere begins to break up into separate spicules. The spicules themselves cannot be regarded as hydrogen filaments, since if we set $\omega_s = 1$ in (34) we obtain $h_m = 4.3$, which is inconsistent with the observations. Hydrogen filaments must have cross sections of only one third of a spicule or less.

The nature of the hydrogen excitation is determined from the ratios of the numbers of hydrogen atoms in the various energy states. The Balmer lines are generally used for such studies, starting with H_γ or H_δ and continuing up to H_8 or H_9, since these lines are all free from self-absorption. If we know the populations for some levels, we can choose values of T_e and n_H for which the observed n_κ agree with the theoretical calculations. On the basis of various observations, we find that for $n_H = 10^{13}$ to 10^{14} and $T_e = 9000$ to 10,000° the theoretical values of n_κ agree with the observed values. These quantities n_H and T_e actually characterize the hydrogen filaments. The main excitation mechanism for the Balmer lines in hydrogen filaments is scattering of the photospheric radiation, as is also confirmed by a study of the profiles of these lines. However, some observations indicate that this mechanism is to a considerable degree supported by excitation due to electron impact and recombination /32/. Finally, it appears that the electron temperature of the hydrogen filaments increases with an increase in solar activity.

The characteristics of the other types of chromospheric filaments may be determined analogously. For helium filaments, we use the line intensities of the helium triplets to compute the populations of the corresponding levels, which are then compared with the populations obtained when the system of stationarity equations is solved. Table 9 gives the relative line intensities for helium, both calculated theoretically and obtained from observations.

The observed level populations appear to be in good agreement with the theoretical calculations for $n_e = 10^9$ to 10^{10} and $T_e = 20{,}000$ to 30,000°, which represent the mean characteristics of helium filaments. The principal mechanism for the excitation of helium triplets is an excitation of the 2^3P level by electron impact, followed by the absorption of photospheric radiation at the frequencies of the corresponding lines.

TABLE 9

	Theoretical	Extra-eclipse observations (averages)	According to Vyazanitsyn	According to Menzel and Cillié	According to Menzel and Athay
$\frac{I4471}{I_{D_3}}$	0.10	0.13	0.079	0.071	0.085
$\frac{I3889}{I_{D_3}}$	0.09	0.10	—	—	—
$\frac{I4713}{I_{D_3}}$	0.009	—	0.009	0.010	0.006

§26. Chromospheric spicules. Photographs of the chromosphere taken during total solar eclipses or during extra-eclipse observations using a polarizing interference filter, as well as spectrohelioscopic and spectroheliographic observations, all show clearly the jagged structure of the chromosphere, which some observers have compared to a prairie fire. Roberts called the individual projections of the chromosphere "spicules." According to many observers, the width of most spicules ranges from 1 to 2", while the average height of a spicule above the solar limb is about 10" No spicules are observed below 6". According to Roberts /33/, spicules are short-lived formations, their average lifetime being about two minutes. Some spicules are parallel to the continuation of the solar radius, whereas some are oriented at a certain angle ($\sim 20°$) to it; the inclination of these does not vary much over large parts of the solar disk. There is a difference between polar and equatorial spicules, and this is more pronounced during years of minimum solar activity. According to Secchi, the spicules are predominantly inclined poleward, but Bugoslavskaya has found that in quiet regions they are mainly directed equatorward. Their inclination may correspond to the inclination of the coronal formations, which may be due to local magnetic fields.

The first data on the spectra of chromospheric spicules were obtained by Michard /34/, who determined the average line-of-sight velocity of the spicules from the displacement of the brightness centers of nodes in the upper chromosphere. The profiles of spicules in H_α were investigated by Smith /35/. According to his data, spicules were observed up to $h = 12{,}000$ km and had a very large Doppler halfwidth, ranging from 1.2 to 1.9 Å. The observed profiles could only with great difficulty be described by the ordinary Doppler formulas. Smith's conclusions were confirmed by Athay /36/. Spicule spectra were also studied at Pulkovo Observatory in 1956, 1958, and 1959 /37/. In these observations, the spectrograph slit was oriented tangent to the solar disk. The spectrograms of highest quality were then processed, which corresponded to typical chromospheric formations. The spicule spectrum was photographed simultaneously in H_α and in D_3; Figures 40 and 41 show the profiles of the H_α and D_3 lines and the H and K lines of Ca^+ in the spicules. Tables 10 and 11 give the equivalent widths (w), the halfwidths ($\Delta'\lambda$), the radial velocities (v_r), and the absolute line intensities (I) for H_α and D_3 and for H and K. In all cases the profiles of the spicular lines are not Gaussian, indicating that the physical conditions in spicules are different from those in the chromosphere. Therefore, velocities recorded using a motion-picture camera should be regarded with

some reservation. These may not be the displacement velocities but rather may be rates of growth or decay of the formation. The Doppler shifts of spicules are such that they almost never reach beyond the limits of the core of the Fraunhofer H_α line for scattered sky light. The average radial velocity obtained for all the spicules measured at Pulkovo had a magnitude of 3.6 km/sec. From the observed third-level populations of hydrogen atoms and from the population of the 3^3D level of helium, it is possible to determine T_e and n_e in the spicules. It appears that n_e is about 10^{10} to 10^{11}, while the concentration of hydrogen atoms is about 10^{10}. Electron-temperature calculations for the helium component of the spicules give possible values of T_e = 20,000°, while for the hydrogen component $T_e \sim 10{,}000°$. All these conclusions are highly tentative, due to the insufficiency of observational data. However, it may be assumed that the spicules represent a phenomenon which is a kind of chromospheric flare, and that they form as a result of the condensation of coronal matter.

TABLE 10

Date	v_r, km/sec	H_α			D_3		
		$\Delta'\lambda$, Å	$w \cdot 10^4$, Å	$I \cdot 10^{-3}$ CGS	$\Delta'\lambda$, Å	$w \cdot 10^4$, Å	$I \cdot 10^{-3}$ CGS
11 May 1958	+2.3	1.93	112	39	0.48	17	60
	+4.6	0.97	355	102	0.47	34	119
	+4.1	1.64	254	70	0.50	36	123
	−1.8	1.20	282	78	0.48	26	91
	+5.5	1.64	210	58	0.47	28	98
	−2.3	1.55	272	75	0.55	29	102
	−4.1	1.61	224	62	0.54	30	105
	+2.3	1.49	320	88	0.54	26	91
	0.0	1.40	78	22	0.60	34	119
	−36.5	0.55	24	7	—	—	—
	−5.5	1.81	410	117	0.62	43	158
	+4.6	1.60	220	61	0.53	33	116
	0.0	1.59	220	61	0.64	27	94
29 September 1959	0.0	0.87	310	85	0.49	16	56
	+7.8	1.03	560	160	0.59	31	109
	+2.3	1.73	287	79	0.78	29	102
	−9.1	1.42	700	200	0.84	22	77
	−4.1	2.38	427	122	0.91	52	182
	0.0	1.04	426	122	0.80	24	84
	−4.6	1.37	260	72	0.59	16	56
	+4.6	1.52	252	70	0.58	32	112
	+1.4	1.24	216	60	0.75	27	94
	0.0	1.28	200	55	0.83	14	49
	0.0	1.66	237	65	0.73	8	28
	−1.4	1.32	60	17	—	—	—
	−4.6	1.25	115	32	0.64	10	35
	0.0	0.89	149	41	0.97	11	38
	−1.4	1.04	311	86	0.62	14	49
	+2.3	1.16	351	97	1.31	51	178
	0.0	1.02	328	90	—	—	—
	1.8	1.05	372	106	0.56	20	70
	−1.4	1.42	368	105	0.94	40	140
	−0.9	1.04	278	77	0.64	16	56
	−0.9	0.94	458	131	0.73	20	70

TABLE 11

v_r, km/sec	K			H		
	$\Delta'\lambda$, Å	$w \cdot 10^4$, Å	$I \cdot 10^{-3}$ CGS	$\Delta'\lambda$, Å	$w \cdot 10^4$, Å	$I \cdot 10^{-3}$ CGS
−3.8	0.38	13.5	4.8	0.40	6.0	2.6
−4.6	0.33	10.7	4.6	0.34	5.2	2.2
+25.8	0.34	7.0	3.0	0.34	3.5	1.5
−1.5	0.34	2.8	1.2	—	1.3	0.6

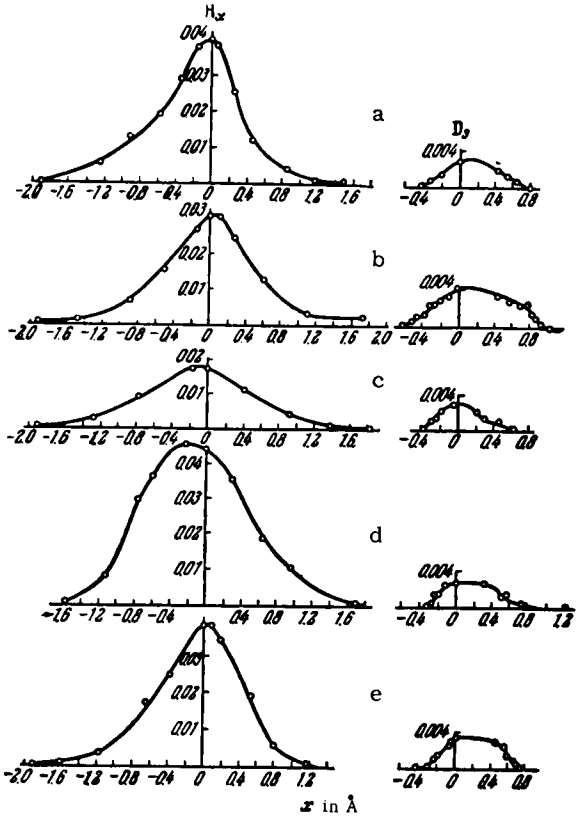


FIGURE 40

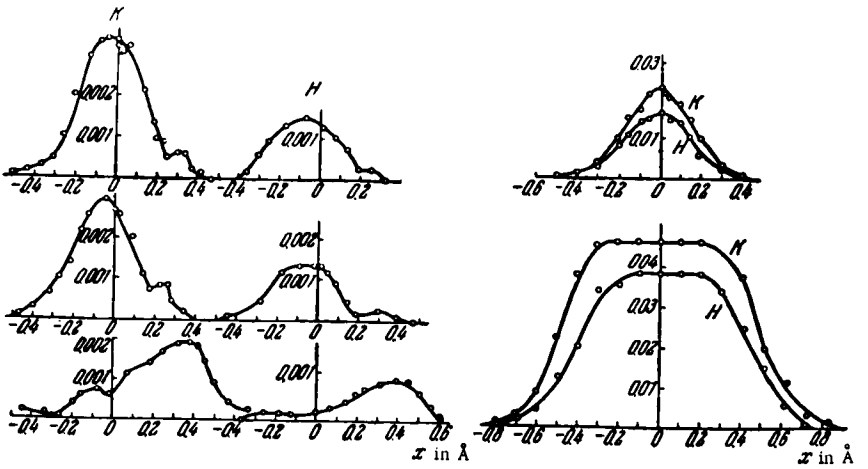


FIGURE 41

§27. The ultraviolet radiation of the chromosphere. The far-ultraviolet region of the chromosphere, which consists of both a continuum and emission lines, is the main factor responsible for the formation of the ionospheric layers in the earth's atmosphere. Even before rocket data on the far-ultraviolet region were obtained, Woolley and Allen /38/ computed the chromospheric and coronal emission on the basis of certain assumptions concerning the nature of the chromosphere. Table 12 gives the results of these computations [as quoted by van de Hulst in "The Sun," edited by Kuiper].

TABLE 12

Emission	Quanta/cm^2·sec ($\times 10^{11}$)	Radiation flux, erg/cm^2·sec (at solar surface)
Continuous emission, lower chromosphere (blackbody, $\lambda < 912$ Å)	1400	3000
Continuous emission, upper chromosphere	260	1000
Line emission, upper chromosphere, $\lambda > 912$ Å	3000	6000
The same, $\lambda < 912$ Å	3000	10000

Starting in 1954, rocket observations of the far-ultraviolet spectrum were made; in 1958 and 1959 data were obtained for the spectrum from 1215 to 83 Å /39/. The most numerous measurements are those of the intensity of the L_α line of hydrogen. Although there are discrepancies between the different observations, due to errors and also possibly due to actual changes in the intensity of the L_α emission as a function of solar activity, still the mean intensity of the L_α emission of the sun is now known quite definitely to be about 3 erg/cm^2·sec (near the boundary of the earth's atmosphere). The other strong line in the far ultraviolet, the He II λ 304 line, was found by Ivanov-Kholodnyi and Nikol'skii /40/ during their analysis of the spectroscopic data published in /39/ to have an average intensity of 1.2 erg/cm^2·sec at the boundary of the earth's atmosphere. In /39/ a list of the lines in the ultraviolet spectrum obtained during rocket measurements is given, together with visual estimates of the relative intensities of the various lines. On the basis of these data Ivanov-Kholodnyi and Nikol'skii constructed a model of the transition region between the chromosphere and the corona, which in their opinion represents the source of the observed ultraviolet spectrum. They showed that the computations of Woolley and Allen were wrong, since they were based on an inadequate chromospheric model due to the lack of the required data on the ultraviolet spectrum of the sun at that time. Ivanov-Kholodnyi and Nikol'skii computed the theoretical intensities of various lines in the ultraviolet region, assuming that the ionization and excitation are mainly due to electron impact; thus, they obtained the distributions of the ions of various elements, according to the degrees of ionization, as functions of temperature. Their estimates are consistent with the observational data obtained using rockets and also with radio-emission data, although the calculations were based on excessively rough values for the effective ionization and excitation cross sections.

From the point of view of the theory of an inhomogeneous chromosphere, the chromospheric emission in the far ultraviolet can be explained by the fact that the ions responsible for the observed lines emit partly in helium filaments (Al II, Mg II, Al III) but mostly in subcoronal chromospheric filaments (there may be several kinds of the latter). The main mechanism of the excitation is electron impact. The L_{α} radiation of hydrogen should originate in both the hydrogen and helium filaments /32/; in the hydrogen filaments the L_{α} radiation is due to scattering of the radiation set up by the weak facular field which virtually covers the entire solar surface. By taking into account that the hydrogen filaments cover 50% of the total area, due to their preferred normal orientation, we can roughly estimate the total line-of-sight intensity of L_{α}. The distance traversed by a ray along the normal to the surface is

$$L = 2\omega_{2,0} \int_0^{\infty} e^{-\beta_2 h}\, dh = \frac{\omega_{2,0}}{\beta_2} = 0.4 \cdot 10^8 \text{ cm},$$

and the population of the second level is $3.5 \cdot 10^4$. Thus, in the absence of self-absorption,

$$I_{L_{\alpha}} = 8.5 \cdot 10^8 \text{ (CGS units)}.$$

However, due to the strong self-absorption, it follows from the theory of growth curves that the intensity will increase in proportion to $\sqrt{N_H}$ and not in proportion to N_H, where N_H is the number of atoms in the line of sight. Consequently,

$$I_{L\alpha} = 3 \cdot 10^4 \text{ CGS units},$$

and since radiation is emitted by only 50% of the hydrogen filaments, we have

$$I_{L_{\alpha}} = 1.5 \cdot 10^4 \text{ erg/cm}^2 \cdot \text{sec}.$$

In the helium filaments the second-level hydrogen atoms originate because of electron impact and recombination, and $n_2 = 10^4$ for $T_e = 25{,}000°$. The distance traversed through the helium filaments is

$$L = 2\omega_{3,0} \int_0^{\infty} e^{-\beta_2 h}\, dh = 1.2 \cdot 10^8.$$

The line-of-sight intensity of the L_{α} radiation of the helium filaments, taking into account self-absorption, is $2.5 \cdot 10^4$ CGS units. Thus, the total emission intensity of all the filaments is

$$I_{L_{\alpha}} = 4 \cdot 10^4 \text{ erg/cm}^2 \cdot \text{sec}.$$

Consequently, at the boundary of the earth's atmosphere the corresponding intensity will be 0.2 erg/cm^2·sec, a value ten times lower than that observed, since the theoretical calculations do not allow for the presence of the facular fields which generate most of the L_{α} radiation.

At present, observational data for the far ultraviolet of the solar spectrum are few and insufficiently accurate, so that they are not yet suitable to

render more precise our knowledge of the physical conditions in the chromosphere.

§28. **The theory of chromospheric heating.** With the discovery of the "anomalous" distribution of the atoms of different elements over the height of the chromosphere, a distribution which did not correspond to the equation of gravitational equilibrium for the solar surface (where $g = 2.74 \cdot 10^4$ cm·sec^{-2} and $T_e \sim 6000°$), the problem arose of the nature of the forces "supporting" the chromosphere. Various attempts have been made to solve this problem during the last decades.

Saha's suggestion (1920) that selective radiation pressure may be an important supporting factor was used as the basis of the first theory, developed by Milne in 1924 to 1928, as applied to the calcium chromosphere /41/. Chromospheric atoms absorbing radiation coming from below must acquire a certain outward momentum $\left(\frac{h\nu}{c}\right)$, where the selectivity of the radiation pressure will correspond to the observed difference in the heights of different elements. This hypothesis seemed particularly likely for the calcium chromosphere, since the resonance lines of ionized calcium are not far from the maximum of the energy-distribution curve of the solar spectrum, while in the chromosphere most of the calcium atoms are ionized. Assuming monochromatic radiation equilibrium in the H and K lines, Milne derived a theoretical density distribution for calcium, as a function of height, which fitted the observations. However, in the 1930's Milne's theory underwent a thorough revision and criticism from both the observational and theoretical viewpoints, and it was shown to be inconsistent in its interpretation of the chromosphere and the prominences.

According to Rosseland's theory (1933), the chromosphere support and the formation of the corona are due to a continual ejection of electrons from the sun, the initial electron velocities being close to the velocity of light. The effect which the resulting positive charge of the sun (set up during the emission of these electrons) has on the state of the atmospheric envelope depends on the degree of its ionization. If the ionization is low, a rapid vertical motion is imparted to the ions, while the greater part of the envelope remains in equilibrium. For total ionization the electric forces act on the entire envelope, which expands until it reaches a state in which the constant flux of ions offsets the loss of electrons. Rosseland considered an atmosphere in which neutral atoms and electrons are maintained in hydrostatic equilibrium, while the ions participate in an outward diffusive motion. By integrating the equations of hydrodynamic equilibrium for these three types of particles, Rosseland found a density distribution which agreed with the chromospheric observations (one or several exponential terms were required). This theory, which was developed only very tentatively by the author, was later abandoned.

After the failure of the theory of selective radiation pressure, McCrea /42/ proposed a theory which explained the low density gradient of the chromosphere as being due to random macroscopic motions, or turbulence. Spectroscopic determinations of the mean turbulent velocity in the chromosphere were made repeatedly, and the results of different authors are fairly consistent ($\sim$ 15 km/sec for the line-of-sight component). Consequently, the thermal component is negligible in comparison with the turbulent component for all elements except hydrogen and helium. For T = 6000° the

hydrogen gradient α is $1.8 \cdot 10^{-8}$ cm^{-1}, which is approximately one third of the purely thermal gradient and agrees more or less with the observations.

Some general considerations on the origin of turbulence in the solar atmosphere were later advanced by Severnyi (1947). Distinguishing between turbulent motions of different scales, Severnyi maintained that energy of "large-scale" turbulence (of stellar size) is continually converted into energy of "smaller-scale" turbulence. The large-scale turbulence may originate due to the decay of laminar rotational motion in the stellar interior, when the angular rotation velocity increases quite rapidly as the axis of rotation is approached.

The conclusion that high kinetic temperatures exist in the chromosphere has had an important influence on the development of the theory. This high kinetic temperature, which is much higher than the temperature of the radiation field, is apparently connected with the existence of a mechanical source which supplies the atmosphere with energy. Therefore, the properties of the atmosphere are mainly determined by the field of the mechanical-energy transfer, rather than by the radiation field, which is incidentally the reason for the observed deviations from conclusions based on "equilibrium" models. This proposition is now adopted as a basic premise in all theoretical works dealing with the explanation of the high chromospheric and coronal temperatures. However, with respect to a specific interpretation of the source and mechanism of the energy transfer, a somewhat different point of view is required.

Thomas /43/ has suggested that the energy is supplied by the ejection of matter from the photosphere in the form of streams moving at superthermic velocities, whose directed mechanical energy becomes partly converted into thermal energy as they move through the atmosphere. He identifies these superthermic streams of matter with the chromospheric ejections or spicules, whose visibility is due to their density being higher than that of the surroundings. According to Thomas's calculations for definite (albeit very hypothetical) values of the parameters involved (the initial velocities of the spicules, their density, etc.), the total energy of the spicular system is 10^{32} erg/sec, and about 10% of this energy is used to heat the chromosphere; such a quantity of heat would be quite sufficient to compensate for the radiation of the chromosphere and to maintain its kinetic temperature at about 35,000°. A similar point of view was expressed previously by V. A. Krat.

M. Schwarzschild /44/ considered the mechanism responsible for maintaining the high temperatures in the chromosphere and corona to be a flux of pressure waves (sound waves) caused when the motion of solar granules from the convective zone is braked by the overlying stable photospheric layers. Such a braking process would transfer mechanical energy to the upper atmospheric layers (designated in general as the "corona") and would impart it to the coronal matter via dissipation. The energy transferred by the granules to the boundary of the convective zone is, according to Schwarzschild, equal to 10^{30} erg/sec, a value which is much higher than the heat discharge of the corona by radiation, * the latter being $6 \cdot 10^{25}$ erg/sec. The efficiency of the energy dissipation depends on the ratio of the material velocity w in the wave to the sound velocity V in the corresponding layer. Consequently, when the mean noise velocity $\bar{w}$ reaches

* The coronal emission is due to free-free electron transitions in proton fields. Electron scattering obviously does not enter into the heat balance of the corona.

a certain critical value $\overline{w}_{cr}$ defined as

$$\overline{w}_{cr}^2 = \beta V^2 = \beta \frac{dp}{d\varrho} = \beta \frac{k}{m} T$$

(where p is pressure, ϱ is density, and $\beta < 1$ is a constant coefficient), even the strongest waves will decay rapidly. At the top of the granulation layer, according to Schwarzschild, $\overline{w} \approx 0.1$ km/sec and $V = 7$ km/sec, so that the waves propagate upward without dissipation and the energy flux $\varrho \overline{w}^2 V$ (and thus also the temperature) remains constant in the lower layers. However, as the density decreases in the outward direction, $\overline{w}$ increases until at a certain height (arbitrarily termed the "base of the corona") it reaches 2 km/sec and dissipation sets in. Between the top of the granulation layer and the base of the corona $\overline{w}$ increases by a factor of about 20, so that ϱ decreases by a factor of about 400; according to the hydrostatic law this corresponds to a layer 800 km thick. This layer includes all the photospheric layers, so we may conclude that the noise flux from the granules transfers its energy through this region without affecting the dynamic or thermodynamic equilibrium there, in agreement with the observations. The noise energy becomes dissipated only in the chromosphere and in the corona, and thus it serves as a heat source for these layers.

An analogous theory was developed independently, and in greater detail, by Biermann /45/, on the basis of the results of theoretical hydrodynamics describing the oscillations of an isothermal atmosphere whose density decreases exponentially in the outward direction. Pressure perturbations with frequencies $\omega \ll \omega_0$ (where ω_0 is the natural frequency of the atmosphere), set up in the photosphere by the pressure of the granules, will cause oscillations only in the part of the atmosphere adjoining the boundary of the convective zone, or else they will produce sustained oscillations of the entire atmosphere (standing waves). Such oscillations are not very effective as agents of energy transfer from the convective zone to the upper atmospheric layers, which can be provided by traveling waves only ($\omega > \omega_0$). During the outward propagation of sound waves the attenuation due to friction and heat conduction is insignificant, and the decrease in the density ϱ is offset by an increase $\Delta v \sim \varrho^{-1/2}$ in all the velocity amplitudes. Consequently, the outward energy flux remains, to a first approximation, constant as long as $\Delta v \ll V$. When Δv becomes equal to V (the case of finite amplitude), then shock waves leading to turbulence are set up, and the energy (whose further transport is rendered impossible) is subject to intense dissipation due to friction and heat conduction. Because of the exponential decrease in density, the dissipation rises greatly toward the outer layers. The physical state of the outer layers of the sun is determined by the fact that the energy brought upward by the pressure waves and dissipated there (its quantity can be estimated from observations of the turbulent velocities) is greater than the energy absorbed from the photospheric radiation; consequently, it follows that the temperature of these layers increases. For the chromosphere this temperature rise amounts to several thousand degrees, if we assume the radiation to be confined mainly to the Lyman continuum.

Although Biermann applies his theory to explain various principal features of the chromosphere (such as turbulence, overexcitation, etc.), he, unlike

Schwarzschild, does not maintain that these same considerations can be extended to the corona. Although from a purely energetic standpoint the application of this mechanism to the corona seems quite probable, still certain facts (turbulent velocities which do not even approximately explain the density gradient or the discontinuous change in the spectrum upon transition from chromosphere to corona, etc.) indicate that the conditions here are radically different and that entirely different interactions must be considered. Moreover, difficulties arise even in connection with the application of the theory to the chromosphere: just like any theory in which the energy is supplied primarily in the form of energy of collisions, this theory is inconsistent with the fact that the chromospheric spectrum corresponds better to radiation excitation, although the necessity for a large energy supply to the upper layers is obvious. The theory was developed further in this direction by Schatzman (1949) and Miyamoto (1949).

Mustel' /46/ divides the chromosphere into two regions according to the height distribution of elements: the lower chromosphere (up to 5000 or 6000 km), where all the chromospheric elements occur (in particular, relatively heavy elements such as Ti, Mg, Se, Sr, and O), and the upper chromosphere (above 6000 km), where only H, He, and Ca occur. In Mustel's opinion, a probable explanation for the greater heights and lower gradients of H and He is to be found in the high kinetic temperature of the chromosphere, which according to Redman is $\sim 30{,}000°$ (hydrostatic equilibrium). The presence of calcium at heights up to 14,000 km and the low calcium gradient are explained by Milne's theory, since the equality of the selective radiation pressure and gravity in the upper chromospheric layers could hardly be accidental. With regard to the other elements (the lower chromosphere), their anomalously low gradients can be attributed to the overwhelming abundance of hydrogen, taken in conjunction with the continuous ejection of matter into the chromosphere from the underlying layers. The moving ejected matter is retarded, scattered, and then becomes diffused (with the exception of H) back downward. Under the effect of gravity only, an atom would fall from a height of 6000 km to the photosphere in 200 seconds, and no very high chromosphere could form; however, the abundance of hydrogen slows down this fall greatly. For example, the equation for the rate of diffusion of gases gives this falling time as 10^8 sec ≈ 3 years.

The works of Alfvén /47/ also deserve mention. This author discusses the heating of the chromosphere and corona by magnetohydrodynamic waves generated by the motion of the granules in a magnetic field. According to Alfvén's calculations, a field of 25 gauss is sufficient to give rise to such waves, which would then travel up into the upper atmospheric layers. Dissipation occurs mainly due to heat conduction (Joule heat); it increases proportionally to $P^{-1/4}$ (where P is pressure) and is the strongest somewhere in the chromosphere or inner corona. These waves cannot exist separately from sound or shock waves.

Recently, Dubov /48/ developed a theory of chromospheric heating by compression waves (similar to sound waves), these being generated in the turbulent region connected with the convective zone of the sun. Part of the energy of the convective elements in this zone is converted into energy of random turbulent motion, which gives rise to compressions and rarefactions (that is, it generates acoustic-type waves which, due to the disordered nature of the turbulence, constitute acoustic noise). These waves propagate

upward and travel through the photosphere without being absorbed. As they progress, the sound waves become transformed into weak shock waves. Then, having reached a low-density level (namely the chromospheric layers), these waves begin to dissipate and their energy is converted to heat which becomes absorbed by the chromosphere and the corona. At certain temperatures and densities a balance will be reached between the energy released and the energy removed by radiation. This balance will be provided by different deexcitation mechanisms at different heights. The entire structure of the chromosphere will be determined by the amplitude of the energy flux transferred by the sound (shock) waves. It is just these shock waves which will be responsible for the formation of regions of different temperature in the chromosphere.

§29. Bright flocculi. The chromospheric formations also include bright calcium and hydrogen flocculi, sometimes referred to as chromospheric faculae. The strongest flocculi occur in active regions. These formations are actually parts of the chromosphere where the self-radiation of the gas above the photosphere is increased. The calcium flocculi on spectrograms are "mottled," without any filamentary formations, whereas the hydrogen flocculi have a distinct vortical structure, especially near sunspots. Figure 42 shows spectroheliograms in H_α and K of Ca II. Considerable difficulty in obtaining good definition is encountered during studies of the physical properties of hydrogen flocculi, since their H_α and H_β emission shows up just as an elevation (and, for H_γ and H_δ, even as a depression) of the central parts of the corresponding lines, which vanishes entirely at the wings of the profile. Therefore, it is very difficult to distinguish the self-radiation of the flocculi from the total radiation emitted by the lower layers of the solar atmosphere. Moreover, there are data which show that the emission of the hydrogen flocculi originates at various levels, that is, that H_α emission is received from layers outside of those emitting H_γ and H_δ.

Mustel' /49/ has discussed in detail the vertical localization of floccular emission in various Balmer lines, during his study of the degree to which the radiation excess at the center of the H lines is due to the chromosphere. He compared the number of excessive quanta emitted at the center of an H_α line with the number of quanta emitted directly by the chromosphere, and he concluded that the chromosphere may be responsible for an emission excess comparable to that obtained from the observations. The H_α emission of the flocculi is thus entirely due to the chromosphere, while for other lines, starting with H_γ, the situation is different. The role of the chromosphere in the formation of the H_γ, H_δ, and H_ε lines is negligible, and the emission in these lines is due to the outer photospheric layers. Since the effective level of floccular emission is located at different heights for different Balmer lines, therefore the emission mechanism may differ correspondingly for the different Balmer lines. Two mechanisms may be indicated as responsible for the excitation of the floccular emission in H_α and H_β: 1) electron impacts and 2) recombination. After studying these mechanisms and taking into account that the electron temperature in the flocculi is less than 7000 to 8000°, Mustel' /49/ concluded that recombination plays the main part in excitation. Calculations show that for $T_e < 7000°$ the recombination processes alone are sufficient to provide the H_α emission of the flocculi. For $T_e = 7500°$, electron-impact excitation becomes as important as recombination.

On the basis of more reliable data on the hydrogen-level populations, V. A. Krat and T. V. Krat /37/ computed the line-intensity ratios for H_α, H_β, and H_γ and compared the results with observations. By applying to these lines the formulas for a quasihomogeneous two-component floccular model, they found that the Balmer decrement of the flocculi should be very steep. The intensity of the flocculi in H_β is therefore low, while in H_γ no flocculi are observed.

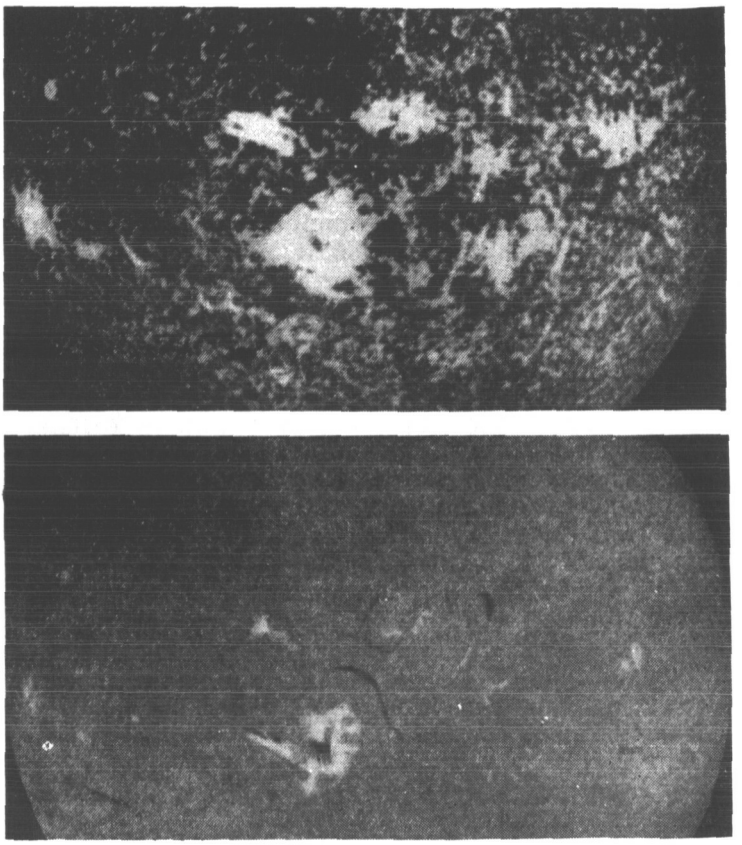


FIGURE 42

A detailed study of the calcium flocculi was made by Mustel' /50/, whose calculations show that the excitation of the H and K lines of Ca^+ is due to electron impact, the combination being negligible. If the calcium and hydrogen flocculi are assumed to emit at the same spot, then the H line in the flocculi (in the case of recombination excitation) would be 30 times stronger than the K line, which does not correspond to the observations. Moreover, computations show that the number of quanta emitted in the H and K lines are considerably greater than the number expected in the case of recombination alone. Finally, by comparing the theoretical line

intensities of the λ 3737 and λ 3706 lines of Ca^+, on the one hand, with those of the H and K lines of Ca^+, on the other hand, assuming a pure recombination mechanism, we concluded that the λ 3706 emission line should be one fourth as strong as the H line, whereas in reality no emission is observed at λ 3706 even in cases of very strong H and K lines. Consequently, the emission mechanisms for the calcium and hydrogen flocculi are different, which is in full agreement with the overall difference in the general appearance of the calcium and hydrogen spectroheliograms.

REFERENCES FOR CHAPTER IV

1. MENZEL, D.H. — Lick. Obs. Publ., Vol.17: 1. 1931.
2. MITCHELL, S.A. —Ap.J., Vol.105: 1. 1945.
3. WILDT, R. — Ap. J., Vol.105: 36. 1947.
4. CILLIÉ, G.G and D.H.MENZEL, — Harvard Obs. Circ. No.410. 1935.
5. PANNEKOEK, A. and M.G.MINNAERT. — Verh. Akad. Amsterdam, Vol.13, No.5. 1928.
6. VYAZANITSYN, V.P. — Izvestiya Glavnoi astronomicheskoi observatorii, Nos. 156 and 160. 1957.
7. KRAT, V.A. — Izvestiya Glavnoi astronomicheskoi observatorii, No. 147. 1951.
8. LINDBLAD, B. and H.KRISTENSEN. — Stockholm Obs. Nederlandes, No.77. 1953.
9. ATHAY, G.R., D.E.BILLINGS, J.W.EVANS, and W.O.ROBERTS. — Ap.J., Vol.120: 94. 1954.
10. HOUTGAST, J. and C.ZWAAN. — Convegno Volta, p.68. 1953.
11. KRAT, V.A. and T.V.KRAT. — Izvestiya Glavnoi astronomicheskoi observatorii, No.153. 1956.
12. KRAT, V.A. and T.V.KRAT. — Izvestiya Glavnoi astronomicheskoi observatorii, No.155. 1956.
13. UNSÖLD, A. — Ap. J., Vol.69: 209. 1929; Zs f Ap. Vol.3: 77. 1931.
14. PEREPELKIN, E.Ya and O.A.MEL'NIKOV. — Izvestiya Glavnoi astronomicheskoi observatorii, No.122. 1935.
15. VYAZANITSYN, V.P. — Tsirkulyary Glavnoi astronomicheskoi observatorii, No.16. 1935.
16. MITCHELL, S.A. and E.T.WILLIAMS. —Ap.J., Vol.77: 1. 1933.
17. MITCHELL, S.A. Eclipses of the Sun. — New York. 1951.
18. WOOLLEY, R. v.d.R. and C.W.ALLEN. — M.N., Vol.108: 292. 1948.
19. MINNAERT, M., G.F.W.MULDERS, and J.HOUTGAST. Photometric Atlas of the Solar Spectrum. — Amsterdam. 1940.
20. PANNEKOEK, A. — B.A.N., Vol.4: 263. 1928.
21. HELD, E.F.M. v.d. — Zs. Physik, Vol.70: 508. 1931.
22. CILLIÉ, G.G. — M.N., Vol.92: 823. 1932.
23. REDMAN, R.O. — M.N., Vol.102: 134. 1942.
24. REDMAN, R.O. — Convegno Volta, p.72. 1953.
25. INGLIS, D.K. and E.TELLER. — Ap.J., Vol.90: 439. 1939.
26. GIOVANELLI, R.G. — Australian J. Sci. Research, Vol.1: 275, 289, 305, 306. 1948.
27. HAGEN, J.P. — Special Suppl. to Journ. Atmos. Terr. Phys. p.353. 1956.
28. WOLTJER, L. — B.A.N., Vol.12: 165. 1954.
29. ATHAY, R.G. and D.H.MENZEL. — Ap.J., Vol.123: 285. 1956.
30. SHKLOVSKII, I.S. and E.V.KONONOVICH. — Astronomicheskii zhurnal, Vol.35: 37. 1958.
31. KRAT, V.A., T.V.KRAT, and L.M.PRAVDYUK. — Izvestiya Glavnoi astronomicheskoi observatorii, 4 (156): 60. 1956.
32. KRAT, V.A. and V.M.SOBOLEV. — Izvestiya Glavnoi astronomicheskoi observatorii, No.160: 116. 1957.
33. ROBERTS, W.O. — Ap.J., Vol.101: 136. 1945.
34. MICHARD, R. — Contrib. Inst. Astrophys., Paris, Ser. A. No. 189. 1955.
35. SMITH, E. — Ap.J., Vol.126: 529. 1957.
36. ATHAY, R.G. — Ann. Astrophys, Vol.21: 98. 1958.
37. KRAT, V.A. and T.V.KRAT. — Izvestiya Glavnoi astronomicheskoi observatorii, No. 167: 6. 1961.
38. WOOLLEY, R. v.d.R. and C.W.ALLEN. — M.N. R.A.S., Vol.110: 358. 1950.
39. VIOLETT, T. and W.A.RENSE. — Ap.J., Vol.130: 954. 1959.
40. IVANOV-KHOLODNYI, G.S. and G.M.NIKOL'SKII. — Astronomicheskii zhurnal, Vol.38(1): 45. 1961.

41. MILNE, E.A. — M.N., Vol. 84: 354. 1924; Vol. 85: 111, 1924; Vol. 86: 8. 1925; Vol. 86: 459, 528. 1926; Vol. 88: 188, 493. 1928.
42. McCREA, W.H. — M.N., Vol. 89: 483. 718. 1929.
43. THOMAS, R.N. — Ap.J., Vol. 108: 130, 142. 1948; Vol. 109: 480. 1949; Vol. 111: 1. 1956.
44. SCHWARZSCHILD, M. — Ap. J., Vol. 107: 1. 1948.
45. BIERMANN, L. and P. TEN BRUGGENCATE. — Veröff. Göttingen, No. 83. 1947; BIERMANN, L. — Naturwiss., Vol. 33: 118. 1946.
46. MUSTEL', E.R. — Izvestiya Krymskoi astrofizicheskoi observatorii, Vol. 2: 98. 1948.
47. ALFVÉN, L. — M.N., Vol. 107: 211. 1947; Cosmical Electrodynamics, — Oxford. 1950.
48. DUBOV, E.E. — Izvestiya Krymskoi astrofizicheskoi observatorii, Vol. 15: 121. 1955.
49. MUSTEL', E.R. — Izvestiya Krymskoi astrofizicheskoi observatorii, Vol. 11: 102. 1954.
50. MUSTEL', E.R. — Izvestiya Krymskoi astrofizicheskoi observatorii, Vol. 9: 25. 1952.

V. P. VYAZANITSYN and V. M. SOBOLEV

N 67-15489

Chapter V

PROMINENCES

§30. Historical remarks. Prominences were discovered in the previous century, when astronomers first began to make physical studies of solar eclipses. During the eclipse on 8 July 1842 in southern Europe, which was observed by the outstanding astronomers and physicists of that time (such as Arago, Struve, and Airy), prominences were recorded by all the observers and aroused considerable interest. In addition, heated discussions took place concerning the question of whether the prominences (and the corona) belonged to the sun or to the moon, or whether they were just an optical illusion. This lively discussion continued for some years, and it was only after the reappearance of prominences during the eclipse of 28 July 1851 in Sweden and Norway that most astronomers reached the correct conclusion that prominences belong to the sun and are gaseous formations. All doubts were dispelled once and for all when Secchi and de la Rue, during the eclipse in Spain on 18 July 1860, were able to obtain successful photographs of the prominences. In addition to giving valuable new information on prominences, Secchi's photographs enabled him to establish the existence of a layer (later called the chromosphere) which surrounds the entire sun and consists of the same luminous matter as the prominences.

A particularly important stage in the development of studies of prominences was the famous solar eclipse of 18 August 1868, when the spectroscope was first applied successfully to an eclipse observation. All the observers (Jannsen, G. Herschel, Tennant, Pogson, and Rayet) obtained results for the spectrum of the prominences which were on the whole consistent with one another, and it was established that the prominences constitute masses of hot gaseous material, the main component being hydrogen. The bright yellow line in the spectrum of the prominences was subsequently attributed by Lockyer to a new element, as yet unknown on the earth; he called it helium, and in 1895 this element was also discovered on the earth by Ramsay. Then, Jannsen and Lockyer established independently that prominences can be observed at times other than during solar eclipses, and that it is even possible to measure the heights and draw the shapes of prominences by successively displacing the solar image relative to the spectroscope slit. Lockyer also discovered the chromosphere using spectral methods.

In 1869 Huggins proposed and first applied in practice a simpler method for prominence observations, namely the use of a widened spectroscope slit, and this method subsequently became the universally accepted one. Therefore, systematic visual observations of prominences on a large scale, including measurements of their positions, heights, and areas, as well as studies of their forms and motions, date from this period. Large numbers of observations have since been made in Italy, Switzerland, Hungary, England, India, China, and other countries.

With the further development of photography, more refined techniques have been devised for the photographic observation of prominences. The scope of spectrographic prominence studies was increased considerably with the advent of the spectroheliograph, a device which not only made it possible to obtain larger and sharper photographs of prominences in the solar-limb emission but also enabled scientists to observe them on the disk itself in the form of dark absorption filaments. The duration of continuous prominence observations became longer, and limb observations which were only 1 to 2 days long were replaced by combined limb and disk observations over periods of almost 2 weeks. The spectrohelioscope, which was developed by Hale in 1924, represented a valuable addition to the spectroheliograph. In the years after World War I Lyot's polarizing interference filter (PIF) became widely used for monochromatic solar observations, and particularly for observations of prominences. Finally, motion-picture methods have become especially important in prominence studies, chiefly when prominence motions are observed. Such methods were first applied for this purpose at the McMath-Hulbert Observatory (USA) at the end of the 1930's and they have since become very popular. In the Soviet Union, motion-picture photography of solar prominences has been carried out continually at the Crimean Astrophysical Observatory since 1948.

During the study of prominences we encounter a combination of exceedingly complex phenomena having a wide physical scope. For instance, one of the most important features of these formations is their various motions. Prominence studies in the present century are associated to a considerable extent with the name of Pettit, who, during the course of thirty years at the Yerkes and Mount Wilson Observatories, collected a large volume of data on the shapes and motions of prominences at the solar limb and also discovered several new facts and regularities. An important contribution was also made by the McMath-Hulbert Observatory, which was founded in 1937 and which was devoted exclusively to the study of solar prominences during the first years of its operation. Valuable information has been supplied by disk observations. In particular, a relation was established between prominences and bright hydrogen flocculi and sunspots, which has helped considerably to understand the development and nature of these formations.*

§31. **Statistical data.** The average annual number of prominences, or the average total area of the prominences observed daily at the solar limb, varies with the phase of the 11-year solar cycle. Systematic observations and statistical studies of this solar-activity index are at present being made at many observatories.**

TABLE 1

Minimum	1913	122	1923	384	1932	350	1944	274	1954	377
Maximum	1917	1187	1926	1285	1938	1145	1948	2047	1958	—

* A list of references on prominences is also given by Waldmeier /1/.

** The results are regularly published in several periodicals: Astron. Mitt. Eidgen. Sternw., Zürich; Osserv. e Mem., Arcetri-Firenze; Kodaikanal Obs. Bulletin; Solnechnye dannye (bulletin published by Pulkovo Observatory and the Solar-Research Commission of the USSR Academy of Sciences).

The curve giving the number of prominences (or the total prominence area) as a function of time is analogous to the corresponding curve for sunspots (Figure 43). For prominence-area measurements unit area is defined as the product of one degree (heliocentric) along the solar limb times one second (geocentric) perpendicular to the limb. Table 1 gives the yearly average prominence areas for several solar cycles (years of maximum and minimum solar activity). The table shows that the variation of the number of prominences during the course of the 11-year cycle is less pronounced than the variation of the sunspot number: the prominence areas in the epochs of maximum are approximately 4 times greater than those in the epochs of minimum, while the corresponding ratios for sunspot numbers are about 20. The ratio of prominence area to relative sunspot number at the epoch of maximum is about 12, whereas at the epoch of minimum it is about 60.

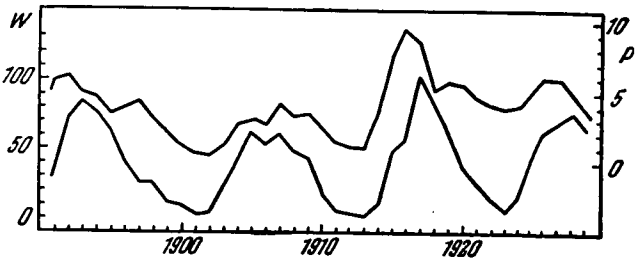


FIGURE 43

Prominences are observed at all heliographic latitudes, but their latitudinal distribution is not uniform. Contrary to the sunspot distribution, it has two maxima in each solar hemisphere. The activity in the low-latitude (or main) prominence zone develops in parallel with the sunspot activity. The average latitude of this zone continually drifts toward the equator from $\varphi \approx 30-40°$, where the prominences appear about 3 years prior to the sunspot maximum or the beginning of the cycle, down to $\varphi \approx 17°$ at the end of the cycle, always remaining about 10° above the sunspot latitude. However, the prominence zone is not demarcated as distinctly as the sunspot zone is, and at the beginning of the cycle it extends from the equator up to $\varphi \approx 60°$. The prominence maximum for the main zone coincides with the sunspot maximum; then, the activity in this main zone diminishes simultaneously with the decrease in sunspot activity, until approximately 1 year before the sunspot minimum the main-zone prominences disappear. The activity in the high-latitude (or polar) zone develops about 3 years after the sunspot maximum, at 45 to 50° latitudes (but not necessarily at the same time in the two solar hemispheres), and it persists at a fairly constant level until the sunspot minimum; the width of the polar zone does not vary much during this period. In the epoch of sunspot minimum, the activity in the polar zone also decreases, and then it starts to increase, at the same time drifting toward higher latitudes; it reaches a maximum about 3 years after the sunspot minimum and then decreases again. In the epoch of sunspot maximum the activity reaches to the poles, after which the prominences disappear temporarily from the high latitudes. This relative phase shift between the activity in the main and the polar

prominence zones reduces somewhat the amplitude of the variation in the total prominence population.

The boundary between the polar and main zones extends approximately along 40° latitude, although a sharp demarcation of the two zones is difficult. The prominence frequencies [of occurrence] at higher and lower latitudes have an approximate ratio of 6:4. A similar latitudinal distribution and time variation of this distribution are observed for the filaments on the solar disk.

According to Waldmeier, in the main zone a fine structure can often be distinguished, exhibiting the presence of a double maximum, the weaker one coinciding with the sunspot zone and the stronger one being displaced about 10 or 15° toward the pole.

The height distribution of the prominences is also of interest. As is to be expected, shorter prominences are more frequent than higher ones. For example, 78% of the prominences have heights of 30 to 60", 16% are 60 to 90", 4% are 90 to 120", and 2% are higher than 120" /2/. The height distribution apparently depends on the heliographic latitude.

§32. **Filaments.** The prominences observed in emission at the solar limb may be detected in absorption at the solar disk. In the disk observations, spectroheliograms using the H and K lines of Ca II or the H_α line show dark elongated formations, called filaments, and occasionally dark amorphous spots as well. Figure 44 gives photographs of two such filaments. In the top row, from left to right, a filament is shown, first photographed in H_α light and in the light of K_3 of Ca II on 13 July 1926 and then photographed on 16 July 1926 in K_3 light. The middle and bottom rows show photographs in K_3 light of a large filament observed moving from the eastern limb (26 August 1929) to the western limb (3 September 1929). The possibility of observing prominences in absorption as well as in emission has widened considerably the scope of studies of these formations.

In the 1930's Newton utilized spectrohelioscopic observations to give a preliminary classification of the dark flocculi, and he established an overall correspondence between the various flocculi shapes and behaviors, on one hand, and the types of limb prominences described by Pettit (see §33), on the other.

Once the height of a prominence at the solar limb is measured, and also the length of the corresponding filament at the central meridian together with its thickness at the point where it appears to be narrowest, all the fundamental dimensions of the prominence can be derived. According to Pettit, the average dimensions of prominences (mostly of quiescent prominences, since these constitute a majority) are as follows: thickness, 10,000 km (range 6000 to 12,000 km); length, 200,000 (range 60,000 to 600,000 km); height, 50,000 km (range up to 100,000 km).

The formation and development of main-zone filaments was studied in detail by d'Azambuja. A filament appears in the active region inside the sunspot zone and then drifts toward higher latitudes, growing on its poleward side and contracting on the side toward the equator, the filament length increasing constantly. Because of the decrease in the angular velocity of solar rotation at higher latitudes, a drifting filament becomes more and more deformed, departing from the meridional direction, until at latitudes above 45° it is oriented along a parallel. Table 2 lists the angle of inclination θ of the filament axis for different latitudes (the averages of

observations at Meudon and Kodaikanal). It is known that when they first appear filaments already show a distortion corresponding to 20 days of meridional deformation due to the differential rotation. Consequently, it may be concluded that about 20 days elapse between the actual formation of a filament and the moment when it becomes visible.

FIGURE 44

TABLE 2

φ	θ	φ	θ	φ	θ
0— 5°	85°	20—25°	44°	40—45°	18°
5—10	82	25—30	43	45—50	16
10—15	71	30—35	34	50—55	16
15—20	60	35—40	26	55—60	9

A filament may remain quiescent for days or weeks, without changing its shape much. Then, however, it may often, without any apparent reason, suddenly become active for a period ranging from a few minutes to many hours, at which time the velocities increase, matter streams into the chromosphere from the prominence, and its shape changes considerably. The prominence finally becomes quiescent again and assumes a shape which is close to its original shape.

A prominence may disappear in various ways: by a gradual waning; by an efflux of matter into sunspots and centers of attraction; or by transition to an active and then an eruptive state, in which the prominence breaks away and escapes into space. The third case was regarded by Waldmeier as being the normal last stage of a large filament (a quiescent prominence). A remarkable feature is that filaments avoid sunspots and faculae; a filament which appears near a sunspot disappears rapidly.

§33. The classification of prominences. The great variety of the outward shapes and the behaviors of prominences, on the one hand, and the accumulation of an enormous amount of observational material on this subject, on the other, naturally led solar scientists to seek a general classification of prominences. On the basis of visual and spectroheliographic observations covering many years, Pettit in the beginning of the 1930's suggested an empirical prominence classification which subsequently became universally accepted. A further development of this classification was made possible by motion-picture observations of prominences. Although at present Pettit's classification appears to have become in part obsolete, still on the whole it agrees with the present-day state of prominence research, particularly from a descriptive point of view.

The entire observable variety of prominences was divided by Pettit into six quite well-differentiated classes /3/: I) active prominences, II) eruptive prominences, III) sunspotprominences, IV) tornado prominences, V) quiescent prominences, and VI) coronal prominences. These basic classes are in turn divided into subclasses. Figure 45 is a schematic diagram of the different prominence types; it also shows the possible relations between classes. The actual forms of the prominences may differ greatly, and only the most general or typical shapes for each case are shown in the diagram.

Pettit's classification is based on the following characteristics of prominences: their relations to sunspots, their origins, their motions, and their structures. With respect to the first characteristic, all prominences fall into two groups: those connected with sunspots and those independent of sunspots (see Figure 45). The latter are in general related to "centers of attraction," which do not show up as any visible formations on the solar surface.

It was previously assumed that all prominences originate in the chromosphere, but we now know that this applies only to certain types, and that some prominences originate in the coronal space or else evolve from prominences of other types. For example, a quiescent prominence may develop into an active prominence, then into a coronal active prominence, and finally into an eruptive prominence. The corresponding evolutionary sequences are indicated in Figure 45 by arrows joining prominences of different kinds.

The types of motion can be as follows: a translational motion of the prominence as a whole; the motion of external jets or streamers (in the form of rays or bands) from the prominence into the chromosphere (toward the center of attraction or toward a sunspot) or into another prominence; motions of knots* along

* A knot refers to prominence material which has a distinct contour, regardless of whether it is located in a streamer or, having separated from the streamer, moves along a trajectory which is free of other luminous matter. On the other hand, a condensation is a bright luminous spot in a prominence or streamer, which differs from the latter only in that it is brighter.

a streamer (or sometimes along a path not followed by a streamer); internal turbulence; and curling into a spiral. Some general kinematic characteristics (downward curling or upward motion) are indicated in Figure 45 by various types of lines (wavy, solid, or dashed) joining the corresponding kinds of prominences. The motions of prominences and their structures are closely linked to one another.

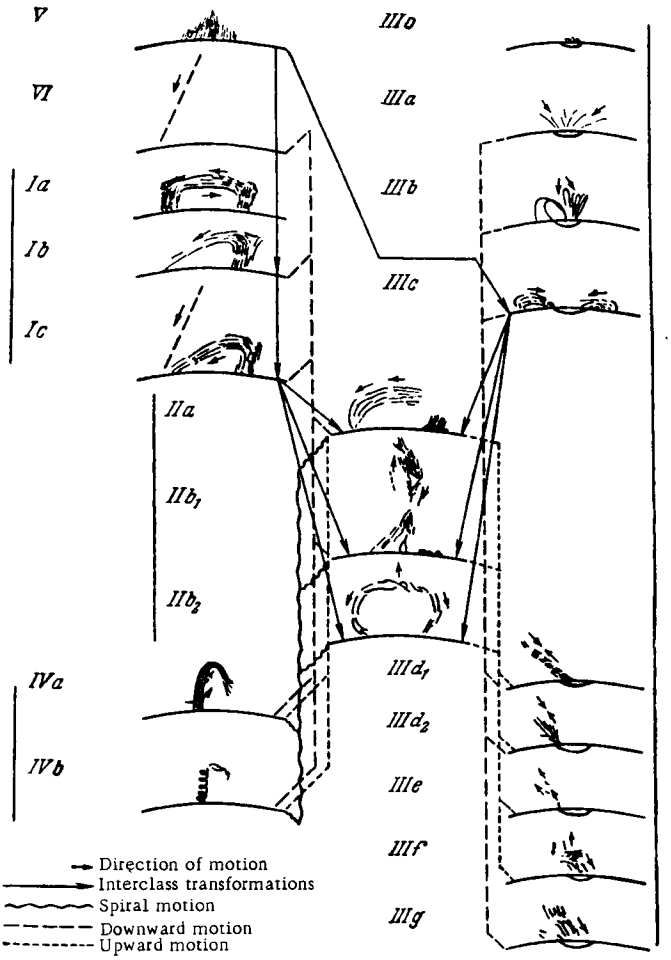


FIGURE 45

Let us now go on to describe the individual classes of prominences. Following Pettit, we will first consider quiescent prominences, since this corresponds to the actual evolutionary sequence of the prominences.

Class V. Quiescent prominences. The name of this class implies low activity and only slight changes in form, position, or structure from one minute to the next. A comparison of spectroheliograms taken at intervals of 4 or 5 minutes with good solar images, and also an examination of motion-picture films, indicates internal turbulence with velocities of

5 to 10 km/sec. A typical feature of this class is the absence of streamers, showing that quiescent prominences are not subject to disturbances by centers of attraction.

Quiescent prominences observed at the solar limb are similar in appearance to haystacks. Their shapes remain virtually unchanged for hours. The best specimens of this class have a kind of palisade structure which is never observed in active prominences (Class I). On the solar disk they often show up as dark wavy filaments. When they form at low latitudes the filaments are oriented parallel to the solar meridians, but as they migrate toward the poles their orientations alter due to solar rotation, until at latitudes above 45° they are directed along the parallels. Thus, prominences at the solar limb near the equator show the greatest number of structural features, while at higher latitudes they become foreshortened and resemble haystacks.

Since prominence activity is a sign of its forthcoming decay, it is no accident that quiescent prominences have longer lifetimes, which may be measured in days or sometimes even in months. Filaments persist on the average through four solar rotations.

Class I. Active prominences. This is the commonest type of prominence. Active prominences can occur at all latitudes from pole to pole, but are found most often in the sunspot zone. Frequently they evolve out of quiescent prominences, and they sometimes develop into the eruptive type.

Active prominences have the form of a mass of interlacing filaments; these filaments appear to be torn apart by a nearby center of attraction. Such prominences are generally joined with the chromosphere and they emit streamers which pour into one or more attraction centers. The luminous material almost invariably separates more readily from the top of a prominence, rather than from its lower parts lying near the center of attraction. This applies to common active prominences (Class Ia).*

Much less frequent are interactive active prominences (Class Ib), that is, two or (occasionally) several prominences joined by streamers via which they exchange luminous material. One of the components may be a tornado prominence. In individual cases of two prominences joined by two streamers, it has been definitely established that the gas moves along the streamers in opposite directions, in other words that the prominences actually exchange luminous material.

The formation of knots and streamers is preceded by increased activity in some part of the prominence, during which several wisps of gas are sent out from the prominence in one direction. Then, the end of a wisp either spearates to form a knot or else elongates into a streamer directed into the chromosphere. Several such streamers may form simultaneously and then converge at approximately the same point in the chromosphere. The diameters of such centers of attraction, which are generally comparable in size to a large sunspot or a small spot group, seldom exceed 3°. The streamers vary greatly in length, ranging from a few thousand kilometers in small active prominences to hundreds of thousands of kilometers in eruptive prominences; the streamer widths are from 500 to 1000 km, and more rarely up to 4000 or 5000 km.

* [In Pettit's system common active prominences are actually Class Ib, while the interactive type constitutes Class Ia.]

Class II. Eruptive prominences. According to Pettit's definition, eruptive prominences are those which rise up from the chromosphere in a more or less vertical direction and spread out into space at enormous altitudes. They do not issue directly from the chromosphere but rather evolve from prominences of Class I or Class III (active sunspot prominences).

The development of an eruptive prominence generally starts with the formation near a quiescent prominence of a center of attraction, into which gas then flows from the quiescent prominence along curved streamers. If the center of attraction is weak, the process may last the whole time required for the entire mass of the prominence to return gradually into the chromosphere. If, however, the force of attraction at the center rapidly increases, then the arc-shaped or arch-shaped streamers moving toward the attractive center will often begin to overshoot it. The prominence then becomes detached and may be rapidly sucked into the center of attraction along the arch-shaped streamers. These so-called quasieruptive prominences (Class IIa) constitute a link between Classes Ib or IIIc and the eruptive prominences proper (Class IIb). If the attraction increases even more rapidly, then the arch-shaped streamers rise until the two branches, ascending and descending, are almost parallel, and the entire prominence starts receding in a more or less vertical direction. An analogous effect is observed if a nearby sunspot is substituted instead of the center of attraction, but this situation is much less frequent. During the course of an eruption a prominence may send out streamers to several centers of attraction (or, which is even more preferable, to a nearby sunspot); in the early stages of the eruption there are quite a number of these streamers, but in the later and final stages only two or three remain.

Two varieties of eruptive prominences are distinguished, although the nature of their motions is essentially the same, In one case (Class IIb_1) a single cloud moves upward, and as it rises it sends down streamers toward the chromosphere. In the second case (Class IIb_2) the prominence is arch-shaped. Here, it is the top of the arch which rises, while the entire arch expands and at the same time gas moves downward along the two side branches.

Different parts of an eruptive prominence may move with different velocities, and at times some parts will have velocities which are radically different from the velocity of the main body. One of the most interesting side phenomena are secondary eruptions, as ejections of isolated gas masses from a prominence are called. These masses move with very high velocities, even higher than the velocity of the main body of the prominence.

Early observers noted that the velocity of vertical motion of an eruptive prominence increases with time or, which is the same thing, with altitude. Later Pettit found that this motion obeys the following two laws: 1) the motion is uniform, but occasionally the velocity may increase in jumps; 2) the new velocity following each jump is a small integral multiple of the preceding velocity, sometimes of the second preceding velocity. *

These laws were derived by Pettit on the basis of graphs showing the measured heights of an eruptive prominence (ordinate) as a function of time (abscissa). In order for the results to be really conclusive, more frequent

* This correction was obtained at the McMath-Hulbert Observatory.

exposures were required, which incidentally was not easy with the older type of spectroheliograph. The advent of motion-picture photography made it possible to reduce the intervals between successive exposures to 1 minute or less. It has been established that the motions of knots and streamers into sunspots and centers of attraction, as well as the motions of quasieruptive prominences, also obey these laws. However, all attempts to explain Pettit's laws have met with failure, and this raises definite doubts as to their actuality.

Pan Puh, who studied this particular problem in 1939, came to the conclusion that these laws were obtained solely because Pettit and his successors applied an inaccurate method based only on the "height-time" graphs.

The usual velocities of eruptive prominences are measured in tens and even hundreds of kilometers per second. A prominence of exceptional velocity was recorded on 17 September 1937 at the McMath-Hulbert Observatory. It first appeared as a Class-IIIc prominence over a large sunspot group in the process of fading out, and it was followed out to an altitude of 1,000,000 km, where it had a velocity of 728 km/sec (a value greater than the escape velocity of the sun), after which it went beyond the field of vision of the camera.

The maximum observed heights for most eruptive prominences are between 100,000 and 500,000 km. The maximum heights observed for prominences were 1,550,000 km, or 1.12 of the solar diameter, on 20 March 1933 near the north solar pole, and 1,703,000 km, or 1.22 of the solar diameter, on 4 June 1946. The lifetimes of eruptive prominences range from a few minutes to 10 hours or longer.

An interesting feature of these prominences is the rapid decay during the last stage of eruption, which generally is a matter of a few minutes; sometimes a very bright prominence will disappear almost instantaneously. In certain cases a prominence will fade out only in part, while another part of it continues to rise. The phenomenon of rapid fading is also observed in quasieruptive prominences as they move toward a center of attraction, as well as in the streamers of active prominences. Extraordinary brightness apparently is not observed in eruptive prominences, at least not unless their velocities are greater than 300 or 400 km/sec. Also, no definite proof has been offered to show that the photosphere or chromosphere increases in brightness near a prominence either prior to or during its eruptive phase.

Eruptive prominences (both Classes IIa and IIb) have a distribution over the heliographic latitude which is similar to Class-I prominences. They have a maximum frequency of occurrence in the sunspot zone, but they are also observed in higher latitudes and even near the solar poles (which are surrounded by centers of attraction). These prominences are one of the less frequent types. Allowing for the randomness of previous observations, the dependence on the season, and other factors, Pettit estimated that the annual number of eruptive prominences on the solar surface is about 400 near the sunspot maximum and 25 to 50 near the minimum.

Class III. Sunspot prominences. This class of prominences is of particular interest, since it gives us information regarding the activity in the chromosphere and the coronal space above a sunspot group or near it. Sunspot prominences are rare, constituting only about 1% of all the prominences. They appear primarily in the early or late development stages of spot groups, and occasionally over a single sunspot. Pettit

divided this class into 9 subclasses. Some of these (a, b, c, and e) are characterized by an influx of gases from the surrounding chromosphere or coronal space into the sunspot region. In other cases (d and f) the material moves out of the sunspot region, returning to the chromosphere along the same route or else escaping from the chromosphere altogether.

IIIo prominences are shaped like chromospheric caps several thousand km high and 20,000 to 40,000 km in diameter, covering a sunspot group. Their lifetimes are usually measured in hours, although at times they persist for a day or more. Large spot groups are sometimes accompanied by two or three such prominences. Many of these prominences are now identified with chromospheric flares.

IIIc (active) sunspot prominences sometimes accompany Class-IIIa or (much less frequently) Class-IIIb prominences. These share many characteristics with Class-I prominences: they send out knots and streamers toward sunspots, they may develop into eruptive prominences, etc.

IIId (returning) sunspot prominences* are temporary chromospheric elevations over sunspot regions, which return to the chromosphere along the same trajectory without any breaks in continuity. They resemble ocean waves breaking onto the shore. They may appear as small patches or as gigantic formations up to 100,000 km or more in size: these may rise and settle without any visible loss of material ($IIId_1$) or else they may disintegrate when they settle ($IIId_2$). The lifetimes are from 10 to 20 minutes for small returning prominences and slightly over an hour for the largest. These prominences may be ejected at any angle to the vertical, and sometimes recur over an active sunspot group, originating at almost a single point.

IIIe (ejection) sunspot prominences are small masses of chromospheric material, sometimes ejected from highly active sunspot regions; these leave the chromosphere and fade out without returning. These prominences are often ejected at large angles to the vertical; sometimes they are spherical (6000 or 7000 km in diameter) and sometimes they look like short streaks. Their velocities are usually higher than the average, and their intensities are about 0.1 of chromospheric, often just at the visibility threshold. Their movements are uniform and rectilinear, but the velocity may decrease suddenly from time to time.

A IIIf (secondary ejection) sunspot prominence constitutes only a very small part of a Class-III prominence. Secondary ejections sometimes take place without disturbing the rest of the prominence. They may move in almost any direction, regardless of the motion of the parent prominence.

Class IIIa, b, and g prominences (coronal sunspot prominences) will be discussed below under Class VI.

Class IV. Tornado prominences. These have the form of spiral columns, often resembling a tightly twisted rope or a finely threaded screw (IVa); they are mostly 5000 to 20,000 km in diameter and 25,000 to 100,000 km high. The presence of rotation has been confirmed spectroscopically; according to Pettit, the rotation is clockwise in both solar hemispheres. A typical feature is the presence of a diffuse smokelike column issuing from the top of the prominence and often inclined to the chromosphere (sometimes even touching it). Unlike terrestrial tornados, tornado prominences have never been found to shift in heliographic latitude; their longitudinal drift is difficult to study because no direct method is

* ["Returning" sunspot prominences (Class IIId) were denoted by Pettit as "surge" sunspot prominences.]

available for the detection of these objects on the solar disk. They apparently decay in various ways: as the peripheral velocity increases, the prominence may explode like a dust devil in the desert; it may rise as a single entity and fade out like an eruptive prominence; or it may evolve into an active prominence.

In addition to columnar tornados, there are also other prominences which appear to consist of individual twisted streamers (IVb). They generally do not have smokelike ejections at the top. Tornado prominences are relatively infrequent.

Class VI. Coronal prominences. These appear as long, slightly curved streamers leading from the coronal space to a center of attraction. Coronal prominences may appear at heights of about 150,000 km and they move downward with constant velocities of about 100 or 200 km/sec; their brightness increases with the approach to the center of attraction. These prominences seldom exist independently, usually being connected with active (Class I) or sunspot (Class III) prominences. The corresponding active or sunspot prominences were classified by Pettit as subdivisions of their groups (Ic, IIIa, IIIb, and IIIg), and these will now be discussed. Coronal active prominences (Ic) appear in the last development stage of a common active prominence, when with the increase in intensity of the attractive center the prominence streamers become more numerous and some coronal streamers may also enter into the center of attraction. Common coronal sunspot prominences (IIIa) are groups of convergent streamers or streamer segments originating in the coronal space over sunspot groups (generally at altitudes below 75,000 km) and directed into the spot region; their velocities are much lower than those of Class-Ic coronal prominences. Sometimes accumulations of streamers in the form of converging arches or closed loops, similar to fountain jets, are observed over active spot groups; these are the so-called looped coronal sunspot prominences (Class IIIb). Sometimes bright points form successively over spot groups, and then gradually combine into a continuous cloud from which streamers pour into the sunspot area. These prominences, known as coronal clouds (Class IIIg), have a tendency to remain at a constant height, but they move parallel to the solar surface in the direction of the main streamers.

Like the prominences of other classes, coronal prominences can be photographed using the lines, for example, of hydrogen and calcium. Their appearance in conjunction with the development of centers of attraction indicates that undisturbed chromospheric gases are present in the coronal space, and that these gases are apparently localized rather than distributed uniformly. Many of the bright coronal prominences constitute chromospheric flares.

The described classes of prominences do not cover the entire gamut of these formations; some other types have been observed but not yet studied properly. For example, although the surface brightness of most prominences in H_α light constitutes 5 to 10% of the chromospheric brightness at the limb, there are still, in contrast to these, some very faint stable prominences, often large and visible with a monochromator, whose brightness constitutes only a few percent relative to the ordinary prominences. "Whirlpool" prominences have also been observed, which have plane rotation and (contrary to Class IV) translational motion. Bright mushroom-shaped

prominences also occur; these may suddenly grow spiral arms which reach out in a definite plane and approach the center of attraction, the gas in them sometimes attaining velocities of 200 km/sec.

Pettit's classification of prominences, despite its grave shortcomings, still to a great extent satisfies the present requirements of observational practice. Its purely empirical nature is an asset, since it may be applied with ease to a descriptive analysis of prominences.

§34. The dynamics of prominences. The motions of prominences are an essential characteristic of these formations. Their external appearances, trends of development, and lifetimes, as well as other features of prominences, all depend on the type and degree of prominence activity, that is, they are closely related to their motions. An understanding of the forces determining the external and internal motions of prominences would therefore be especially helpful in clarifying the nature of these formations. These forces have been studied by many scientists, but they have still not been satisfactorily explained. The very presence of prominences in the coronal space, particularly quiescent prominences which persist for a long time in the solar gravitational field without changing much in appearance, leads to the assumption that certain forces existing in prominences can counteract the force of gravity. The uniform motion of eruptive prominences (see Pettit's first law), as well as other factors, appear to indicate this. It is difficult to doubt that the force "supporting" the chromosphere (see §6), on the one hand, and the force "supporting" the quiescent prominences, on the other hand, are identical in nature and are both electromagnetic forces.

Another interesting fact which the observations have revealed is that the different elements present in prominences are well-mixed, that is, there is no separation of elements. The problem of the separation of elements has a long history. It first arose in connection with the radiation-pressure theory of the chromosphere and the prominences. Various arguments were presented in favor of the separation of elements: the previously overemphasized division of prominences into ordinary and metallic, the different radial velocities of prominences and the different chromospheric heights for different lines, etc. These arguments, however, were later shown to be untenable. For instance, Pettit's comparison of the spectra of ordinary and metallic prominences taken during a solar eclipse using a flash spectrum showed that all three of the formations studied had the same chemical composition, and that the differences in total brightness and in the relative line intensities were due to differences in the optical thickness and in the degrees of excitation. Numerous comparisons of the prominence forms using light of various wavelengths did not reveal any real difference: under conditions of identical photographic densities of the images using the various lines, the forms of prominences (that is, down to the streamers and the fine structural features) remain the same, showing that the elements are not separated.

On the basis of motion-picture recordings of prominence motions, taken in H_α hydrogen light using polarizing interference filters, Severnyi and Khokhlova /4/ concluded that all prominences can be divided into three classes according to their motion and variation of form:

1) eruptive (ejection of the prominence or a considerable part thereof from the sun);

2) electromagnetic (active), in which regular motions are observed along well-defined trajectories;

3) quiescent, which are characterized by random motions.

The prominences in Classes 2 and 3 are the most frequent, eruptive prominences being the rarest. The latter generally are formed when quiescent prominences develop. The development commences with an acceleration of the gas to high radial velocities, mainly at the base of the prominence. A large part of the prominence then starts rising, first slowly and then rapidly; different parts of the prominence move almost uniformly at different velocities, often experiencing sharp changes (mostly increments) in velocity, which may sometimes be as high as 500 to 700 km/sec. When the ascending mass of the prominence reaches a considerable height above the surface of the sun, knots, and sometimes streams of gas as well, separate from it and move back toward the sun. The abrupt velocity variations observed in eruptive prominences are substantial, and they constitute a distinctive feature of this class (the so-called first law of Pettit). Eruptive prominences represent a special case of the rapid, explosive development of quiescent prominences.

Electromagnetic prominences are characterized by perfectly ordered, directional motions along definite, usually curved, trajectories, with knots and jets of gas passing frequently along these trajectories. Unlike eruptive prominences, this type is not characterized by discontinuous velocity variations, and the knots and jets move with constant velocity over the major part of the trajectory (that is, Pettit's laws are not observed). The velocities are some tens of kilometers per second.

The direction of motion in an electromagnetic prominence depends as a rule on the positions of the "centers of attraction," these generally being localized at the solar surface and at the surface of the interacting mass of the prominence. The mass of the prominence may lie either in the chromosphere or in the corona. Therefore, in addition to upward motion (from the chromosphere to the "center of attraction") along an arched trajectory, there also exist downward motions (from the corona to the "center of attraction").

The main properties of the motions of these prominences imply that the motions are related to the presence of magnetic fields in the solar atmosphere. The trajectory shapes are very similar to the magnetic lines of force around a pole or dipole. Also, the constant-velocity motion along curved trajectories can be attributed to the motion of charged particles in a constant magnetic field.

During studies of the motions in quiescent prominences it would be more accurate to speak of changes rather than motions. Prominences of this type have various, sometimes strange, forms. The changes in them are slow, and take hours or days; they consist in random motions of individual knots, clouds, and jets, as well as changes in the form of the structure as a whole, somewhat reminiscent of the changes of shape in terrestrial clouds. Quantitative measurements of the motions in quiescent prominences are very difficult. A comparison of the observed distribution of knots (according to velocity) with the Maxwellian distribution fails to reveal any similarity. This noncorrespondence between the velocity distribution function of the knots and the normal function of statistical physics shows that the phenomenon involves turbulent motions.

A study of the brightness fluctuations of quiescent prominences, carried out by Dubov /5/ using motion-picture films taken in the H_α line, made it possible to determine some of the turbulence parameters. It was found that

in many cases the knots of a quiescent prominence can be treated as density fluctuations randomly scattered throughout a field of isotropic turbulence, the dissipation energy being around $5 \cdot 10^8$ erg/deg·sec. The scale of turbulence is approximately 10^9 cm and the turbulent velocities are about 10 km/sec.

§35. Prominence spectra. Spectrophotometric measurements. The spectra of prominences, as well as their motions, are of basic importance in a study of the physical nature of prominences. The spectrum of a prominence observed at the solar limb (that is, in projection against the sky) is an emission-line spectrum. The strongest lines are the first members of the Balmer hydrogen series, the H and K lines of ionized calcium, and the D_3 line of helium. During a total solar eclipse, when the scattered sunlight (both the sky background and scattering in the instruments) is reduced to a minimum, many of the weaker lines become observable: these include about 40 Balmer lines and 26 Paschen lines of hydrogen, a few lines of neutral helium and the He^+ 4686 line, and the lines of neutral and ionized metals (see Table 3).

TABLE 3

Element, λ, Å		Element, λ, Å		Element, λ, Å		Element, λ, Å	
Ti^+	3322.91	Mn^+	3482.95	Fe	4383.54	Fe^+, Fe	5168.99
Fe^+, Fe	3341.88	Ti^+	3685.25	Ti^+, V	4395.13	Mg	5172.65
Ti^+	3349.00	Ti^+	3759.33	Ti^+	4443.85	Mg	5183.58
Ti^+	3349.41	Ti^+	3761.33	He	4471.54	He	5875.64
Ti^+Sc^+	3361.24	Ti^+	3761.88	Ti^+, Fe^+	4504.03	Na	5889.98
Ti^+	3372.84	He	3819.63	Ti^+, Fe^+	4549.63	Na	5895.99
Ti^+	3383.84	Mg	3838.30	Ti^+	4563.76	He	6678.15
Ti^+, Fe	3394.56	Fe	3859.87	Ti^+	4572.00	He	7065.18
Cr^+	3421.25	Ca^+	3933.90	Fe^+	4583.86	Ca^+	8498.06
Cr^+, Fe	3422.71	Ca^+	3968.70	He^+	4695.63	Ca^+	8542.13
Fe	3440.62	He	4026.28	He	4713.15	Ca^+	8662.17
Fe	3441.01	Sr^+	4077.83	He	5015.68	Mg	8807
Mn^+	3441.97	Sr^+	4215.70	Fe^+	5018.44		—
Mn^+	3474.11	Ca	4226.74	Mg	5167.35		—

Beyond the Balmer series (λ 3646) a weak recombination continuum is observed, whose intensity, in accordance with theory, decreases exponentially toward higher frequencies. In the brightest prominences it is possible to distinguish an even weaker continuum in which the energy is distributed according to frequency as in the solar spectrum; this continuum is due to the scattering of solar radiation by free electrons.

In the spectrum of a quiescent prominence the majority of lines are those of hydrogen and the metals, the latter being mostly ionized. Neutral metals have fewer lines, and these mainly correspond to the strong Fraunhofer lines. The spectrum of an eruptive prominence apparently does not differ from that of a quiescent prominence. Sunspot prominences, on the other hand, typically have spectra in which strong metal lines are present (strontium, magnesium, sodium, titanium, iron, etc; in addition, there are strong helium lines), and this is why they are called metallic prominences. It should be noted that even under extra-eclipse conditions a quite intense continuum can in many cases be observed in metallic prominences. However, the difference between the spectra of quiescent and metallic prominences is undoubtedly due not to a difference in chemical

composition but rather to differences in the excitation conditions and in the optical thickness. A comparison of their spectra, obtained during solar eclipses, shows that the spectrum of a quiescent prominence mainly comprises lines whose intensities on Mitchell's scale for the chromospheric lines are higher than 30, whereas the spectrum of a metallic prominence also contains weaker lines, with intensities less than 15. The difference between the spectral lines of the two types of prominences very often reduces to a difference in the intensities, and this can naturally be interpreted as due to a difference in the degrees of excitation and ionization or a difference in the amount of emitting atoms, the actual chemical composition being the same.

Comprehensive investigations of prominence spectra have been undertaken by Vyazanitsyn /6/, Zuikov /7/, Severnyi /8/, Ivanov-Kholodnyi /9/, Zirin /10/, Sobolev /11/, Tandberg-Hansen and Zirin /12, 13/, and Yeh Shih-Hui /14/. Consequently, a fairly large amount of spectrophotometric material has been accumulated, from which it was possible to obtain information on the physical characteristics of prominences. Of greatest interest are spectra including wide wavelength ranges, obtained using spectrographs with quite high dispersions and resolving powers. All the authors have observed that the profiles of the hydrogen and helium lines in prominences are in the majority of cases Gaussian. Self-absorption is often significant, not only for the first Balmer lines H_α and H_β but also for the lines up to H_ε /14/. Figures 46a and 46b show the profiles of some of the lines in prominence spectra. Using the profile halfwidths for various lines, it is possible to obtain information on the kinetic temperature (the actual temperature of the gas), the turbulent velocity, and the populations of various levels.

One of the methods for determining the kinetic temperature is based on the familiar formula

$$\frac{\Delta\lambda_D}{\lambda} = \frac{1}{c}\sqrt{\frac{2RT_{kin}}{\mu} + v_t^2}. \qquad (1)$$

Here $\Delta\lambda_D$ is the Doppler halfwidth of the line, which can be obtained from observations using the formula

$$\Delta\lambda_D = \frac{\Delta\lambda}{2\sqrt{\ln 2}}$$

($\Delta\lambda$ being the total profile width at half the central intensity), μ is the atomic weight, R is the gas constant, and v_t is the turbulent velocity.

Formula (1) is valid provided the profile is produced solely by the Doppler effect, which many observations have shown to be the case in prominences. If we assume that the turbulent velocities of the atoms of hydrogen, helium, and the metals are all equal, then we can determine the kinetic temperature and the turbulent velocity by writing this equation for hydrogen and the metals or for helium and the metals and solving for the two unknowns (T_{kin} and v_t). Vyazanitsyn /6/ has shown that the turbulent velocities of hydrogen atoms and calcium ions are approximately the same. This, however, cannot be taken as definitely the case for all elements. Moreover, this method is applicable only if hydrogen and the metals or helium and the metals are excited in the very same regions, a condition

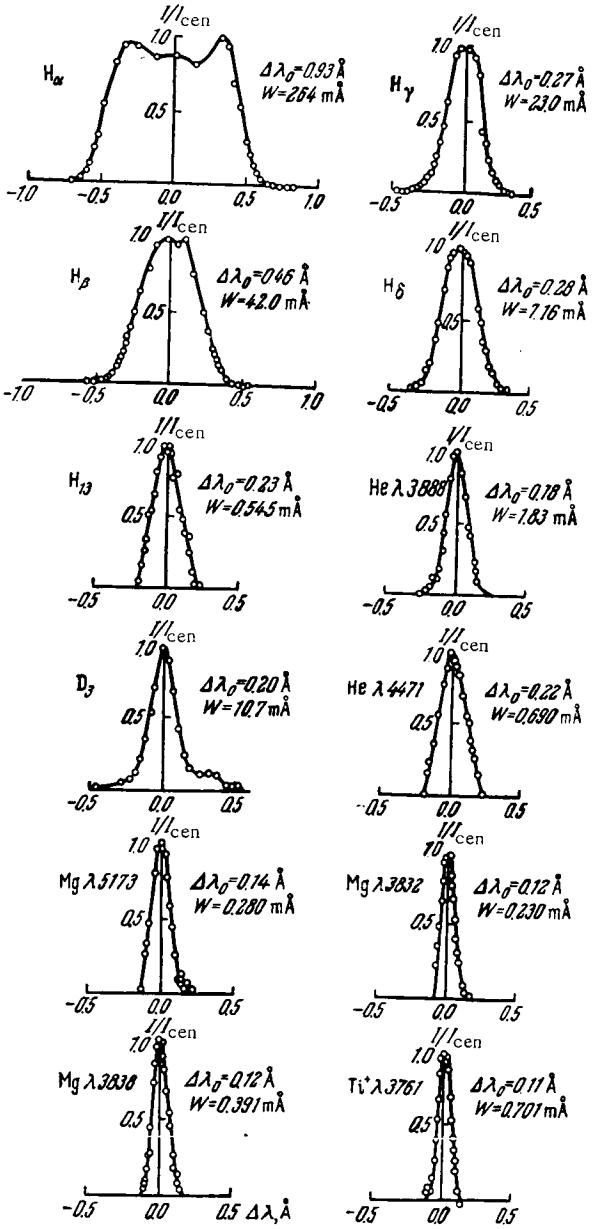

I/I cen
Hα
Δλ0=0.93 Å
W=264 mÅ
Hγ
Δλ0=0.27 Å
W=23.0 mÅ
Hβ
Δλ0=0.46 Å
W=42.0 mÅ
Hδ
Δλ0=0.28 Å
W=7.16 mÅ
H13
Δλ0=0.23 Å
W=0.545 mÅ
He λ3888
Δλ0=0.18 Å
W=1.83 mÅ
D3
Δλ0=0.20 Å
W=10.7 mÅ
He λ4471
Δλ0=0.22 Å
W=0.690 mÅ
Mg λ5173
Δλ0=0.14 Å
W=0.280 mÅ
Mg λ3832
Δλ0=0.12 Å
W=0.230 mÅ
Mg λ3838
Δλ0=0.12 Å
W=0.391 mÅ
Ti+ λ3761
Δλ0=0.11 Å
W=0.701 mÅ
Δλ,Å

FIGURE 46 a

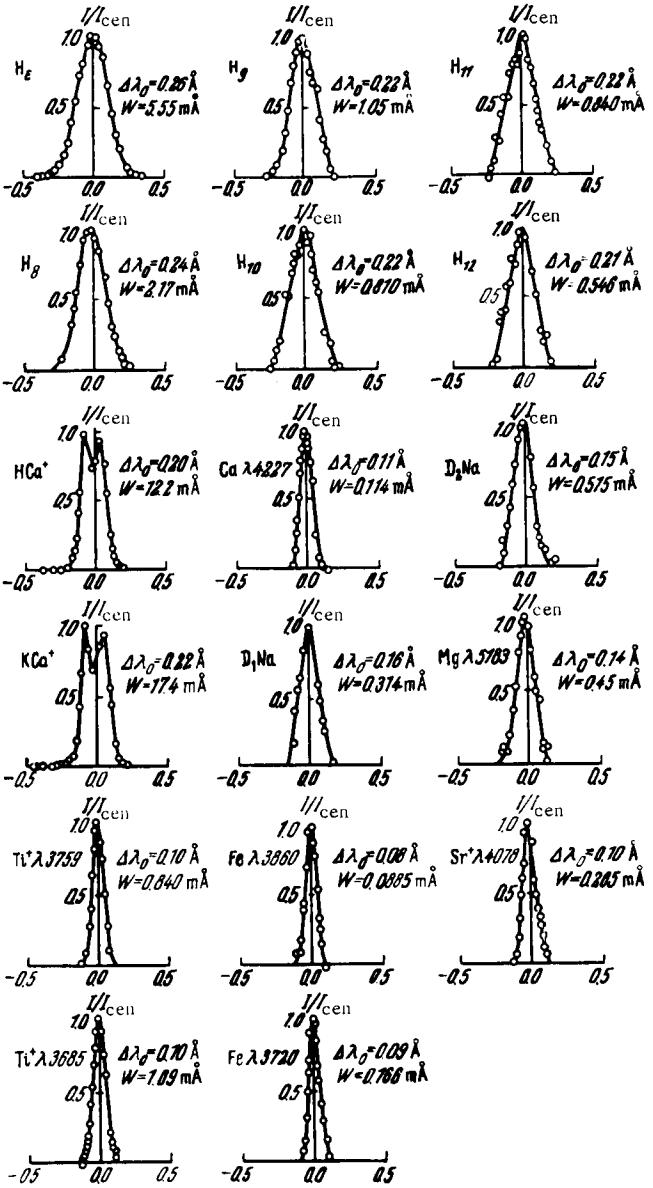

I/Icen
Hε
Δλ0=0.26 Å
W=5.55 mÅ
Hθ
Δλ0=0.22 Å
W=1.05 mÅ
H11
Δλ0=0.22 Å
W=0.940 mÅ
H8
Δλ0=0.24 Å
W=2.17 mÅ
H10
Δλ0=0.22 Å
W=0.810 mÅ
H12
Δλ0=0.21 Å
W=0.546 mÅ
HCa+
Δλ0=0.20 Å
W=12.2 mÅ
Ca λ4227
Δλ0=0.11 Å
W=0.714 mÅ
D2Na
Δλ0=0.15 Å
W=0.575 mÅ
KCa+
Δλ0=0.22 Å
W=17.4 mÅ
D1Na
Δλ0=0.16 Å
W=0.374 mÅ
Mg λ5183
Δλ0=0.14 Å
W=0.45 mÅ
Ti+ λ3759
Δλ0=0.10 Å
W=0.840 mÅ
Fe λ3860
Δλ0=0.08 Å
W=0.0885 mÅ
Sr+ λ4078
Δλ0=0.10 Å
W=0.285 mÅ
Ti+ λ3685
Δλ0=0.10 Å
W=1.09 mÅ
Fe λ3720
Δλ0=0.09 Å
W=0.768 mÅ
1.0
0.5
-0.5
0.0
0.5

FIGURE 46 b

which is not always satisfied. The determinations of T_{kin} and v_t carried out by different authors show large discrepancies, indicating the great variety of physical conditions existing in prominences. As an example, let us consider the values of v_t and T_{kin} obtained by some authors for hydrogen.

According to Vyazanitsyn /6/, the turbulent velocity is 10 km/sec and the kinetic temperature is 5200° ± 2800°; according to Zuikov /7/, v_t = =13.5 km/sec and T_{kin} = 9000°; and Ivanov-Kholodnyi /9/ obtained v_t= 7.1 km/sec and T_{kin} = 11,800°. For different prominences Yeh Shih-Hui obtained values of v_t ranging from 3.76 to 7.06 km/sec and values of T_{kin} from 5120° to 7790°. The quoted values of v_t and T_{kin} are in most cases the results of averaging for various prominences. Yeh Shih-Hui has demonstrated that these parameters differ for different prominences.

From the measured line intensities it is possible to find the numbers of emitting atoms and consequently the chemical composition, the degree of excitation and ionization, and certain other characteristics of the prominences.

Let us consider a column 1 cm^2 in cross section and parallel to the line of sight in a prominence. We denote the energy emitted by this column in a given line ($k \to i$ transition) per sec per unit solid angle toward the observer as I_{ki}; also, we denote the number of emitting atoms in the column as N_k, the Einstein coefficient of spontaneous-transition probability as A_{ki}, Planck's constant as h, and the frequency of the $i \to k$ line as ν. If the optical thickness of the prominence in the given line in the direction specified is much less than 1, then

$$I_{ki} = \frac{N_k A_{ki} h\nu_{ik}}{4\pi}, \qquad (2)$$

since forced transitions for fairly high frequencies are negligible under the conditions prevailing on the sun. Once I_{ki} is determined from observations, we can calculate the value of N_k from (2).

However, in many cases we must take into consideration the self-absorption in the H_α, H_β, and H and K Ca^+ lines. Calculations show that for the ordinary optical thickness of prominences line broadening due to natural damping is negligible, so that we can assume a pure Doppler coefficient of absorption (Vol. I, Chapter 1, §8), the Doppler width being determined by both the thermal and the turbulent motions. The main difficulty entering into the self-absorption problem is that we do not know the function J_ν, which depends on the mode of excitation of the atoms in the chromosphere. In the simplest case (which is the one most important in practice) of the scattering of solar radiation in the prominences, we have $J_\nu = B_\nu$, where B_ν is the average intensity of the incident radiation.

We can obtain an empirical curve of growth by plotting the intensities of some particular line, say the K line of Ca II, measured in various parts of a given prominence (or different prominences), against the corresponding intensities of the D_3 line of He, which is in some cases free from self-absorption. In this case it is also assumed that $\bar{J}_\nu$ is constant for the given prominence (or different prominences); this is equivalent to assuming a universal curve of growth, which is obviously a shortcoming of this method.

Sometimes it is possible to allow for self-absorption with greater accuracy, provided we take advantage of certain specific features of the case being considered. For example, the values of $\bar{J}_\nu$ are approximately equal

for the H and K lines of Ca II, which are components of a multiplet (oscillator-strength ratio of 1:2) and which are located in the same spectral region; thus, the effect which self-absorption has on the ratio $I_H : I_K$ is determined by the quantity N_i alone. On the basis of equation (5) of Chapter IV, we can obtain a theoretical relation between these quantities; here, the velocity of the atoms, which enters into the expression for the absorption coefficient $\varkappa$, can be determined from the halfwidths of the H and K lines. Applying this method, Vyazanitsyn /6/ found, as an average for many prominences, that there were $2.4 \cdot 10^{13}$ atoms of Ca II per cm^2 along the line of sight, or, taking the average thickness of a prominence as 10,000 km, that there were $2.4 \cdot 10^4$ Ca II atoms per cm^3. It was also found that on the average self-absorption reduces the brightness of the H line of Ca II to one third. These studies have established that self-absorption for the H_α line is considerable, that it is especially strong for the H and K lines of Ca II, and that it is negligible for helium lines and the weaker metallic lines.

A problem which is more difficult than the measurement of equivalent widths is the measurement of line profiles. Such measurements have been made during the last three decades by various researchers, in particular by Soviet astronomers, but the data are still meager.

Vyazanitsyn /15/ has measured the profiles of the H and K lines of Ca II in the rare case in which one prominence is shielded by another. Specifically, the spectrogram shows the wide emission lines of the lower, lagging prominence, which contain the narrow absorption lines of the leading prominence at their centers; up above, these absorption lines change into emission lines without any change of width. Assuming pure Doppler broadening, the line profiles yielded the following results: the optical thickness of the leading prominence at the center of the H and K lines (neither of which was exceptionally bright) was ≈ 2.5; the line-of-sight component of the turbulent velocity was 13.6 km/sec; the number of absorbing Ca II atoms per cm^2 parallel to the line of sight was $1.1 \cdot 10^{13}$; the attenuation of lines due to self-absorption was 50%; and the intensity of the exciting radiation was 0.025 of the intensity of the adjacent continuum at the center of the solar disk, that is (allowing for an attenuation to about 1/3 due to dilution), the excitation was due to emission at the center of the Fraunhofer H and K lines, whose central intensity is ~ 0.08. These data are apparently typical for prominences.

§36. Excitation and ionization. Studies of excitation and ionization are complicated considerably by the absence of thermodynamic equilibrium in prominences and by the presence in their spectra of the lines of elements which differ greatly in excitation potential (hydrogen, helium, and metals). In order to solve the problem, we must write and solve for each element a system of stationarity equations, at the same time taking into consideration the various possible elementary processes responsible for the population of levels and for deexcitation. The following processes may cause level excitation:

1) the scattering of photospheric radiation;

2) excitation by electron impact, both from the ground state and from the lower excited states;

3) recombinations at excited levels;

4) the absorption of radiation emitted by the prominences (for resonance lines);

5) spontaneous transitions from upper levels to a given excited level.

Excited states may be destroyed by:

1) spontaneous transitions to lower levels;
2) ionization by electron impact and photoionization;
3) electron impact of the second kind.

The elements in prominences may be ionized by the absorption of radiation (of either internal or external origin) at frequencies of the continuum beyond the main and subordinate series (photoionization) and by electron impact from both the ground state and excited states.

An exact solution of the system of stationarity equations which takes into consideration the aforementioned processes is impossible even for hydrogen and helium, since the information, if at all available, which we have concerning the elementary processes is highly inaccurate (for instance, we cannot specify the effective electron-impact excitation and ionization cross sections, particularly for transitions from excited levels, the radiation densities in resonance lines, etc.).

Therefore our analyses of the excitation and ionization in prominences are still purely qualitative, and are directed toward ascertaining the roles of the various mechanisms. Studies of this subject carried out by several authors /6, 16, 11, 12, 13, 14/ have shown that, due to the great variety of physical conditions in prominences and due to their inhomogeneity, the excitation and ionization may be governed by different mechanisms.

Vyazanitsyn /6/ solved the stationarity equations for hydrogen considering only excitation by photospheric radiation and by recombination; he thus obtained the Balmer decrement as a function of the electron temperature and of the Planck temperature for photospheric radiation and for the Lyman-continuum radiation. He showed that a Lyman-continuum temperature of 6650° and a kinetic temperature of $T_{kin} = 5200°$ suffice to explain the observed hydrogen-level populations in most of the prominences considered.

During a study of the distribution of hydrogen atoms according to different states in quiescent prominences, Ivanov-Kholodnyi /16/ concluded that excitations of the third, and in part of the fourth, levels are caused by scattering of photospheric radiation in the H_α and H_β Fraunhofer lines, whereas the excitation of higher levels is determined by recombination.

However, if recombinations are of any significance in the excitation of hydrogen, then we cannot neglect excitation by electron impact, since a comparison of these mechanisms (see /11/) has shown that electron impact and recombination participate equally in the excitation of hydrogen. Actually these two mechanisms are functionally related and are determined by the same parameters, namely the ground-state concentration of hydrogen atoms and the kinetic temperature. Therefore, a recombination spectrum can appear only in exceptional cases, when the ionization is determined by the field of external coronal radiation, which is possible only for very rarefied formations having comparatively small optical thicknesses at the frequencies of the Lyman continuum.

Sobolev /11/ has used the hypothesis that prominences are inhomogeneous formations consisting, like the chromosphere, of filaments with various physical conditions to demonstrate that the spectra which he obtained for certain prominences could be explained by excitation due to the absorption of photospheric radiation; the excitation in other prominences, on the other hand, was mainly determined by recombination and electron impact. As

an example, Table 4 gives the Balmer decrements (relative to the intensity of the H_β line) calculated theoretically and obtained from observations.

TABLE 4

Spectral line	Electron-impact and recombination excitation				Excitation by photospheric radiation	
	prominence of 10 March 1954	theoretical, for $T_e = 10{,}000°$ and $n_H = 10^{11}-10^{12}$	prominence of 26 May 1954	theoretical, for $T_e = 7\,500°$ and $n_H = 10^{13}$	prominence of 17 May 1954	theoretical, for $T_e = 5\,000°$
H_α		3.9	2.1	2.8	3.1	8.0
H_β	1.00	1.00	1.00	1.00	1.00	1.00
H_γ	0.46	0.46	0.51	0.45	0.30	0.28
H_δ	0.23	0.23	0.32	0.29	0.13	0.11
H_ε	0.15	0.14	0 14	0.16	0.060	0.054
H_8		0.10		0.12	0.061	0.063
H_9		0.07		0.09	0.055	0 050

Different prominences thus correspond to different physical conditions, and consequently to different excitation mechanisms. This conclusion was earlier drawn by Ivanov-Kholodnyi /16/ and later confirmed by other authors.

Tandberg-Hansen and Zirin /12, 13/ made a study of the lines of hydrogen and helium and the λ 4686 line of ionized helium and concluded that these elements are excited in regions with different electron temperatures, so that prominences can be divided into two classes according to the type of excitation. The main characteristics of this division are the intensity ratios for the He I 4026 and Sr II 4077 lines, on one hand, and the He I 4713 and He II 4686 lines, on the other. Table 5 gives some estimates of these ratios.

TABLE 5

Class I	Class II
$I_{4026} \sim I_{4077}$ $I_{4713} \gg I_{4686}$	$I_{4026} \gg I_{4077}$ $I_{4713} \sim I_{4686}$

Prominences of Class I have comparatively strong metallic lines and a very weak He II 4686 line (to use the terminology of Tandberg-Hansen and Zirin, these are "cold" prominences). Class II is characterized by weak metallic lines, whereas the intensity of the He II 4686 line is comparable to that of the He I 4713 line ("hot" prominences).

Yeh Shih-Hui /14/ obtained the spectra of 10 prominences using an echelle spectrograph which made it possible to obtain simultaneously, with a quite high dispersion, the entire visible spectrum and a large part of the ultraviolet region. After determining from the observational material the main physical characteristics for hydrogen (the level populations and the kinetic temperature), he estimated the relative significance of the various processes in the excitation of hydrogen. In addition, he set up stationarity equations (with certain simplifications) for the observed populations and T_{kin} and then determined the ground-state concentration of hydrogen and the electron concentration. Yeh Shih-Hui's main conclusion is that the excitation of hydrogen in prominences is due mainly to their self-radiation.

However, all these studies have not given a comprehensive description of the excitation, and particularly the ionization, in prominences. They have only furnished data on the significance of the various mechanisms in excitation and ionization and have indicated the great variety of physical conditions in prominences.

In order to explain the emission of prominences in the lines of various elements, we must obtain a quite rigorous solution of the system of stationarity equations for all elements whose lines are observed in prominences. Moreover, we must have at our disposal a much greater amount of observational data on the spectra, since in many cases different spectral regions were photographed at different times and we can by no means be sure that they represent the very same part of the prominence, the resulting data being unreliable. The only exceptions to this are Yeh Shih-Hui's observations, but these cover just 10 prominences, mostly of the quiescent type.

The problem of the origin of prominences is of considerable significance in understanding the excitation and ionization processes. One possible mechanism of prominence formation, as a result of the condensation of coronal matter, has been considered by Shklovskii /17/. The main requirement for such a process is the existence of a magnetic field in the form of a system of force tubes, in which the coronal matter becomes compressed. This process is not an adiabatic one, and the thermal energy released during compression will be transmitted to the surroundings by heat conduction and by the radiation emitted in free-free and free-bound transitions. Calculations show that in a few hours the temperature of such a compressing gas would drop to 200,000°, after which the cooling process would become more rapid until in a few minutes the temperature would drop to values of around 10,000°.

REFERENCES FOR CHAPTER V

1. WALDMEIER, M. Ergebnisse und Probleme der Sonnenforschung, 2 Auflage. —Leipzig. 1955.
2. VYAZANITSYN, V.P. —Izvestiya Glavnoi astronomicheskoi observatorii, No.145: 68. 1950.
3. PETTIT, E. —Ap.J., 76:9. 1932; 84:319. 1936; MCMATH, R. and E.PETTIT. — Ap.J. 85:297. 1937; 88:244. 1938.
4. SEVERNYI, A.B. and V.L.KHOKHLOVA. —Izvestiya Krymskoi astrofizicheskoi observatorii, 10: 9. 1953.
5. DUBOV, E.A. —Izvestiya Krymskoi astrofizicheskoi observatorii, 15: 121. 1955.
6. VYAZANITSYN, V.P.— Izvestiya Glavnoi astronomicheskoi observatorii, No.136. 1947.
7. ZUIKOV, V.N. —Izvestiya Glavnoi astronomicheskoi observatorii, No.155. 1959.
8. SEVERNYI, A.B. —Izvestiya Krymskoi astrofizicheskoi observatorii, 12: 33. 1954.
9. IVANOV–KHOLODNYI, G.S. —Izvestiya Krymskoi astrofizicheskoi observatorii, 13: 112. 1955.
10. ZIRIN, H. —Ap.J., 124 (2): 451. 1956.
11. SOBOLEV, V.M. —Izvestiya Glavnoi astronomicheskoi observatorii, No.158: 1960.
12. TANDBERG-HANSEN, E. —Ap.J., 131 (3): 717. 1960.
13. ZIRIN, H. and E.TANDBERG-HANSEN, —Ap.J., 131:717. 1960.
14. YEH SHIH-HUI.— Izvestiya Krymskoi astrofizicheskoi observatorii, 15:180. 1961.
15. VYAZANITSYN, V.P. —Izvestiya Glavnoi astronimicheskoi observatorii, No.139: 31. 1947.
16. IVANOV-KHOLODNYI, G.S. —Izvestiya Krymskoi astrofizicheskoi observatorii, 16: 69. 1955.
17. SHKLOVSKII, I.S. Fizika solnechnoi korony (Physics of the Solar Corona). — Fizmatgiz. 1962.

N 67-15490

V. A. KRAT

Chapter VI

CHROMOSPHERIC FLARES

Chromospheric flares are the brightest formations on the sun, but at the same time are very short-lived; they are visible in the light of the H_α line. It has been established that, in certain ways, flares are very much like explosions. The sudden increase in radiation flux at the location of a flare is similar to phenomena occurring in variable stars, mainly in UV Ceti variables. It is possible that the latter case represents a related phenomenon, although on a completely different scale. Chromospheric flares are also interesting in that they may have strong effects on the ionosphere and the magnetic field of the earth. Moreover, they emit radio waves and may affect the cosmic-ray particles incident upon the earth; in fact, they are sometimes even sources of such particles (of the solar component of cosmic radiation).

§37. **Classification. Frequency of flares.** Until recently the spectrohelioscope was the standard instrument used for flare observations. A flare was identified by means of a sudden rise in brightness in the H_α line at some part of the solar disk. At present, however, spectrohelioscopes are used concurrently with telescopes equipped with polarizing interference light filters, which, having a transmission-band halfwidth of about 0.5 Å, make it possible to view simultaneously the entire solar disk in H_α light. Considerable progress has also been made employing motion-picture methods. Motion pictures make it possible to trace the development of large chromospheric flares and also to carry out the necessary photometric measurements.

The development of a flare, according to observations in H_α, takes place in a complex manner, starting with the appearance of luminous points or bright diffuse cloudlets having highly fluctuating, yet on the average increasing, brightnesses. The rise in brightness for small flares takes no more than 1 or 2 minutes, and it is usually because of this that the commencement of a flare is seldom detected by an observer. On the other hand, the decay stages of such flares may last quite a long time, reaching as much as an hour in individual cases. Visual estimates of the class of a flare are made according to an arbitrary four-point scale.

TABLE 1

Class	Average duration (minutes)	Maximum and minimum durations (minutes)	Average area (10^{-4} of a solar hemisphere)	Average brightness (units of disk brightness in neighboring continuum)
1	17	4— 43	1	0.6
2	29	10— 90	4	1.0
3	62	20—155	10	2.0
3+	180	50—432	—	—

The area and brightness of a flare at the time of its maximum development are taken as characteristics of the flare class. Table 1 gives the average class characteristics according to Ellison /1/ and Waldmeier /2/. Weak chromospheric flares are very frequent. During years of maximum solar activity several flares may often be observed within a single active region (a spot group surrounded by a facular area) in the course of an hour. Class 3 or 3^+ flares, on the other hand, are rare events, and no more than one or two of these occur each month.

FIGURE 47

Chromospheric flares are generally localized in the lower chromospheric strata, together with the bright hydrogen flocculi. Large flares as a rule have complex irregular structures (Figure 47), but sometimes they may resemble vortexes (Figure 48). In certain cases, the bright flocculi of flares have joined spots of opposite magnetic polarity. Some flares are observed at considerable altitudes above the solar limb in the form of anomalously bright prominences.

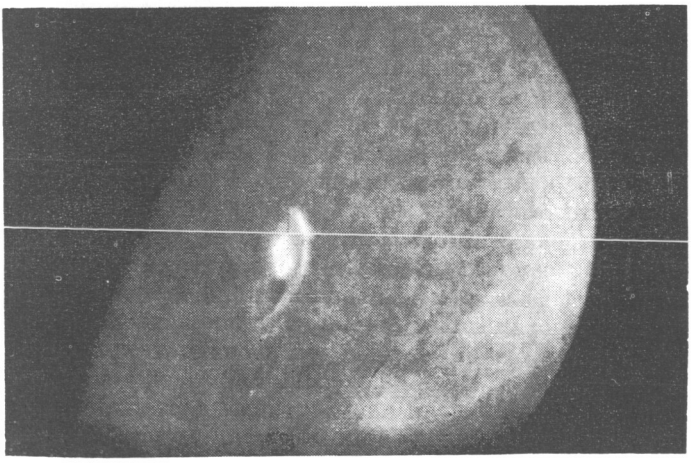


FIGURE 48

The problem of the gas motion in flares cannot be solved conclusively just by observations in the line cores. Many authors have been of the opinion that flares are totally at rest, but this conclusion cannot be accepted, since the halfwidth of the H_α emission line in a flare may reach 1 Å. Consequently line shifts may remain imperceptible even at comparatively high velocities, of the order of 10 or 20 km/sec. In the event of very large shifts, attenuation of the flare will be observed. In spite of the limitations of such observations, Severnyi and Shaposhnikova recently detected in some flares bright formations moving with velocities up to 300 km/sec. Although theoretically these motions could be considered as displacements of the emission center of the gas, still certain indirect factors, and also the results of spectroscopic observations, indicate an actual motion of the gas masses.

Still earlier, at the moment of flare birth, Dodson, Hedeman, and Chamberlain /3/ observed in five cases that individual bright jets issued from the flares, the jet velocities being up to 700 km/sec. The observations of flare spectra (Figure 49) made by Krat and Sobolev /4/ at Pulkovo showed that the emergence of bright jets from flares is not an exception; they were also observed in Class-1 flares. The jets move predominantly downward, toward the photosphere.

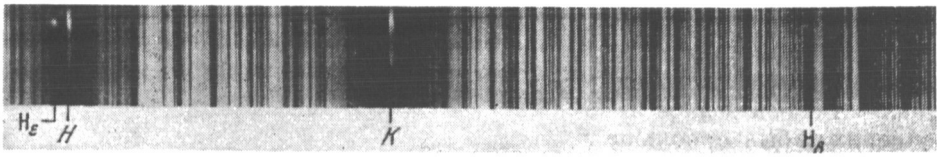


FIGURE 49

§38. **The formation of surges.** Special prominences known as surges also appear during chromospheric flares. On the disk they show up as small, short-lived, dark flocculi, whereas near the limb they appear as bright, rapidly moving prominences. Although surges are observed only comparatively rarely, still we can be sure that they almost always (if not always) accompany flares. However, because of the high radial velocities involved, the spectral lines of surges are shifted as much as 10 Å off the centers of the Fraunhofer lines. Therefore an observer who is occupied with spectrohelioscopic observations of a flare at the center of the H_α line will generally miss a surge, which moreover may even occur comparatively far away from the flare proper. During observations using polarizing interference filters surges often cannot be detected, since it is impossible to recenter the filter pass band during the observations. Thus, the data available on surges are comparatively few.

During spectrohelioscopic observations a surge is generally noticed when its radial velocity reaches several tens of kilometers per second. If a line-shifter is used to move the pass band of the filter from the line center into its violet wing, then a dark blob from 10 to 30" in diameter is observed. The blob mass elongates and its velocity increases. After a few minutes, the velocity decreases and eventually drops to zero, at which time the surge will have reached its maximum height. Then it starts falling downward, following its previous path of ascent. In the long-wave (red) wing of the

line the blob remains distinct until its actual disappearance. The velocity of descent is much lower than the rise velocity was, and it generally does not exceed the free-fall velocity. At high altitudes, surges often break up to form individual jets and blobs (Figure 50).

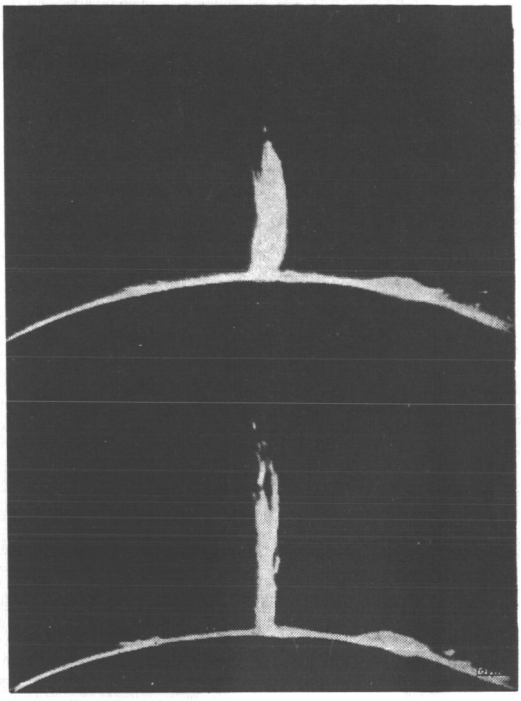


FIGURE 50

Observations of surges at the solar limb show that the surge brightness may be either higher or lower than the disk brightness. When the surge is brighter, then if it is observed on the disk it will inevitably be taken for part of the chromospheric flare. However, the term "chromospheric flare" should apparently just be applied to the virtually stationary nucleus, all the rapidly moving blobs and jets being classified as surges. The recent observations of Teryaeva have shown that the spectra of flares and surges are essentially identical. The most characteristic feature of surge motion is the high amount of velocity dispersion. This can be explained only by assuming that many small blobs and jets, moving at various velocities, are projected simultaneously onto the spectrograph slit. It has also been determined that fairly strong magnetic fields exist near surges (see §40) and that the surges move along the magnetic lines of force.

§39. **The spectra of chromospheric flares and their interpretation.** The spectra of chromospheric flares were first studied systematically by Richardson and Minkowski /5/. These authors made observations using the emission lines of H, He, and Ca II, from He 10938.12 to Ca II 3736.919. Later, Allen /6/ compiled a list of 116 lines of a chromospheric-flare spectrum obtained using a three-prism spectrograph. This list was made on the

basis of spectroscopic and spectrographic (visual) observations of 9 flares. The most complete list of the lines to be observed simultaneously in an intense flare is the one published by Severnyi, Steshenko, and Khokhlova /7/, who observed flares using an echelle spectrograph, which enabled them to photograph on a single plate the spectrum from 3200 to 6600 Å

In summing up the attempts toward a more complete description of the emission spectra of flares, we may state that the spectra of strong flares include all the lines of hydrogen, neutral helium, and ionized calcium and strontium, while in some of the "blobs" all the other lines of the ionized and neutral metals are observed (Mg, Fe, Fe II, Sr II, Cr, Ti II, Ba II, etc.). Thus, practically the same lines are observed as in the spectrum of the chromosphere, except for the ionized-helium line He II 4686, which is extremely rare in flare spectra.

Very comprehensive and reliable spectrophotometric studies of chromospheric flares were made by Mustel' and Severnyi /8/, and later by Severnyi, at the Crimean Astrophysical Observatory. These authors showed that in flares of Classes 1 and 2 noticeable changes are observed only in the H and K lines of ionized calcium and in the Balmer lines of hydrogen. In the H and K lines, the central emission is accompanied by an elevation of the line wings, usually observed not at the commencement of the flare but only after a certain delay. The authors attribute this phenomenon to ionization of Ca II ions by the Lyman emission of the flare. The time lag is due to the gradual diffusion of Lyman quanta into the reversing layer, as a result of the high absorption coefficient at Lyman frequencies. A similar elevation of the line wings is observed in the Balmer lines, and the time lag is apparently due to the same factor. Observations of a large flare (Class 3^+) on 5 August 1949 gave some most remarkable results. The authors obtained 17 spectrograms of the flare during the eight-minute period after its commencement, and thus they were able to pick up the peak of its development. However, no changes in the metallic lines were observed for this flare either. Only the profiles of the H and K lines and the λ 4215.5 ionized-strontium line were found to change. It should be noted, by the way, that other authors have not observed any wing elevation in the Balmer lines. Subsequently, Severnyi established that the extended wings of the Balmer hydrogen lines and the H and K lines are mainly due to the Doppler effect. The wings may often be asymmetric. The line halfwidths in flare spectra are very large, and they are also "Doppler" in nature, corresponding to velocities of some tens of kilometers per second inside the flare. It is just because of this enormous Doppler broadening that the Stark effect in the hydrogen lines is generally obscured.

An interesting phenomenon observed in this flare (and also reported previously) is a depression of the short-wavelength wing of the H line and of other Balmer lines. This indicates the presence of gas efflux with velocities up to 80 km/sec from the region of the flare, which can be interpreted as a surge prominence. One blob in this surge was observed in H_α emission, and it was found to move with a velocity of 100 km/sec. The wide wings of the H_α line were satisfactorily explained as being mainly an effect of radiation damping; the wings of the other Balmer lines are formed partly due to the Stark effect. Thus it was possible to evaluate the physical conditions in those regions of the flare in which hydrogen was excited. The line-of-sight number of excited hydrogen atoms in the second quantum state (N_2) was

found to be between $3 \cdot 10^{15}$ and $1.3 \cdot 10^{16}$, while the electron concentration n_e (the number of free electrons per cm^3) ranged from $3.3 \cdot 10^{12}$ to $1.1 \cdot 10^{13}$. In accordance with the line profile, it was assumed that $T_{kin} = 10{,}000°$ (a gas temperature which is practically equal to the electron temperature under the conditions prevailing in flares). This result was confirmed by Krat and Sobolev /4/ during their study of weak chromospheric flares (Class 1); they showed that the temperature in the hydrogen-emission zone did not exceed 10,000°. The values of N_2 and n_e for these flares are about 10 times less than for the flare of 5 August 1949. This is sufficient to ensure that the Stark effect plays only a negative role in the formation of the profiles even of the H_ε, H_δ, and H_9 lines.

Emission lines of helium (especially the D_3 line) are observed on the disk only in very large chromospheric flares /7/. However, the corresponding absorption lines are often associated with flares /9/ and do not disappear even after the flare is over. Since the D_3 and He 3889 lines are often observed over sunspots and simply in facular regions, however, they need not be attributed to flares.

The main mechanisms responsible for the emission lines of flares are apparently atomic recombination and electron impact. The large populations of the higher hydrogen-atom levels, which are the reason for the many spontaneous transitions from the higher levels to the second level, producing the Balmer series, may be a result of higher temperatures in the hydrogen-emission zones of some flares. T_{kin} for individual flares may be as high as 15,000 or 20,000°.

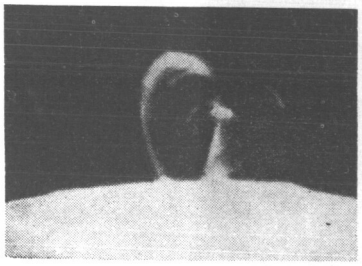


FIGURE 51

Photographs of the spectra of flares located above the solar limb have recently proved to be very fruitful. At the limb flares are generally observed as bright prominences, whose gas escapes into the chromosphere in the form of individual jets. As a result of a misunderstanding, limb flares were classified as sunspot prominences, even when the jets flowing out of the flares formed loops (these were called loop prominences, see Figure 51). Actually, when prominences of these types are projected onto the solar disk, they are absolutely indistinguishable from flares observed in H or K light.

The development of such prominences is identical to the development of flares on the disk. By observing the downward flow of gas jets from these prominences, we can understand the appearance of the "red" (descending) surges which are almost always observed in flare spectra. In some cases flares rising over the photosphere have actually been seen to develop directly into one or several sunspot prominences at the solar limb. In particular, this phenomenon was recently observed by Waldmeier during the magnificent flare of 23 March 1958. This flare also showed intense continuum emission.

The spectrum of a flare observed above the solar limb is remarkable due to the abundance in it of weak lines, most of which cannot be detected in the spectra of flares projected onto the solar disk. Recently, a large number of spectrograms of limb flares have been obtained at the Climax

Observatory (USA) and at the Pulkovo and Crimean Observatories. The spectrograms of the large flare on 24 June 1956, taken by Zirin at the Climax Observatory with an extra-eclipse coronagraph, show the emission lines of He 4922, 4713, 4471, and He II 4686. All the lines are very wide, their halfwidths being on the average 0.8 Å. The halfwidth of H_β was even as great as 2.75 Å. Such large widths are indicative of nonthermal line broadening; in the given case, however, as will be shown in the following, the broadening factor is apparently not turbulence but rather a regular gas motion of a compression or expansion type.

Flare spectra obtained at the Pulkovo Observatory in 1958 reveal a great variety of halfwidths and intensities in the helium lines. It is interesting that He II 4686 is by no means observed in all flares. However, it must be kept in mind that for this line to be detectable its intensity must be at least comparable to that of the weak He 4713 line. In the flare of 24 June 1956 the intensity of the He II 4686 line was even higher than that of He II 4471, which is always very strong in flare spectra.

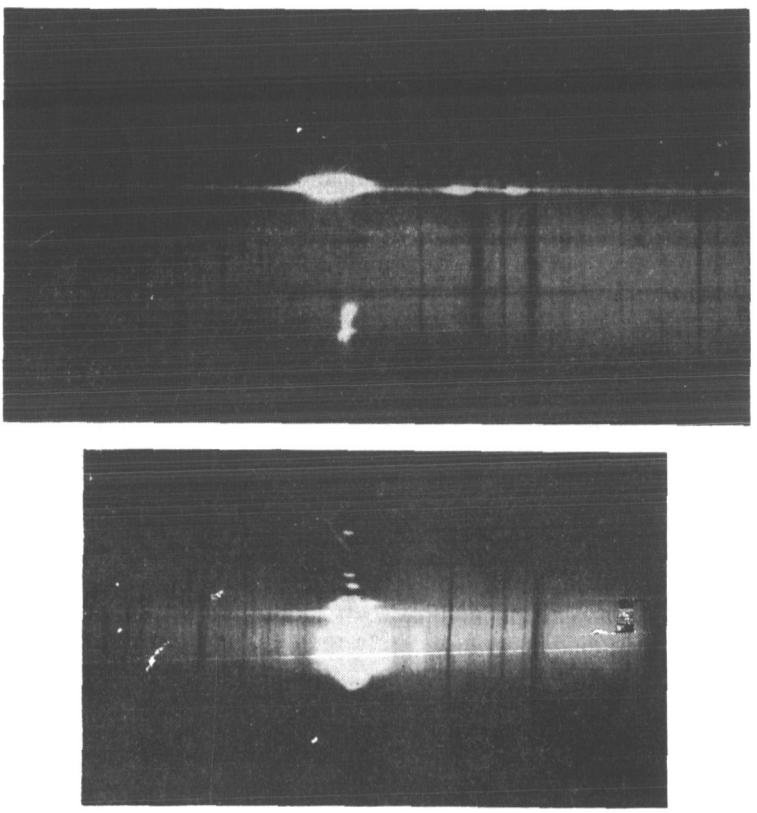


FIGURE 52

The emission lines of metals (Na, Mg, Ti II, Fe II, etc.) are concentrated in a comparatively narrow region near the flare center. The hydrogen-

emission zone, on the other hand, is more extended, and appears to surround on all sides the "metallic" core of the flare. Even more extended than this is the zone of helium emission. The H_γ and H_β emission lines of hydrogen, and also the D_3 emission line of helium, often have unusually wide wings. Photographs taken at Pulkovo by Prokof'eva and Gol'dberg-Rogozinskaya using an extra-eclipse coronagraph show considerable emission, often at very large distances from the line center; this indicates gas motion with velocities of several hundred kilometers per second (Figure 52). On the spectrogram of the flare of 5 April 1959 (bottom photograph) Gol'dberg-Rogozinskaya was able to detect in the long-wavelength wing of the H_α line an emission corresponding to a velocity of 930 km/sec. The Doppler effect is the only plausible explanation of the line broadening in flare spectra.

These features of the behavior of the helium and metal lines indicate not only that the flares themselves are excessively inhomogenous (since the conditions of excitation and ionization differ completely for different elements) but also that the conditions for excitation and ionization of the atoms of different elements in different flares are not the same.

The gas motions in loop prominences (flares) and the spectrophotometric data help us to understand the development of flares. A flare apparently develops during the compression (condensation) of coronal gas, which rapidly cools off as the line emission intensifies. Compression is actually the direct opposite of explosion, although the gas-compression velocities in flares, to judge from the radial velocities in H_α and D_3, are supersonic and thus approach the velocities of detonation waves. Compression explains very well the temperature drops at flare centers and in their dense "metallic" cores, as well as the appearance of intense magnetic fields in flares due to a concentration of magnetic lines of force. Severnyi and his co-workers observed an intensification of the magnetic field up to 600 gauss in the flare of 31 August 1956; the maximum field strength did not coincide in time with the maximum hydrogen emission.

If the magnetic field is intensified during the flare, then obviously this strengthening cannot be caused by the flare itself, since then there would be no reason for compression to occur. The reason for the magnetic-field variation and for the resulting gas compression can only be related somehow to the properties of the magnetic field. In this respect, two possibilities must be considered. The coronal magnetic field may change either due to its instability (more precisely, due to the instability of the coronal plasma) or else due to some extra-coronal factor, such as the magnetic-field variations in sunspot groups. The observations of Gopasyuk indicate that the appearances of flares are accompanied by changes in sunspot groups. The motion of a spot is much more intense than the motions on the quiescent sun, and spot motions are apparently fluctuational-translational in nature.

§40. Emission cores. In 1917 at the Mount Wilson Observatory Ellerman observed some brief flares in the wings of the hydrogen lines. Narrow emission bands were detected in the wings of H_α, while the core of the line remained dark. This indicated that a flarelike phenomenon had taken place in the reversing layer or even in the photosphere of the sun. Ellerman observed that the short-wavelength wing of this emission, which he called a "bomb," is generally longer than the long-wavelength wing. However, not much attention was paid to Ellerman's discovery, and until quite recently these "bombs" were not studied further.

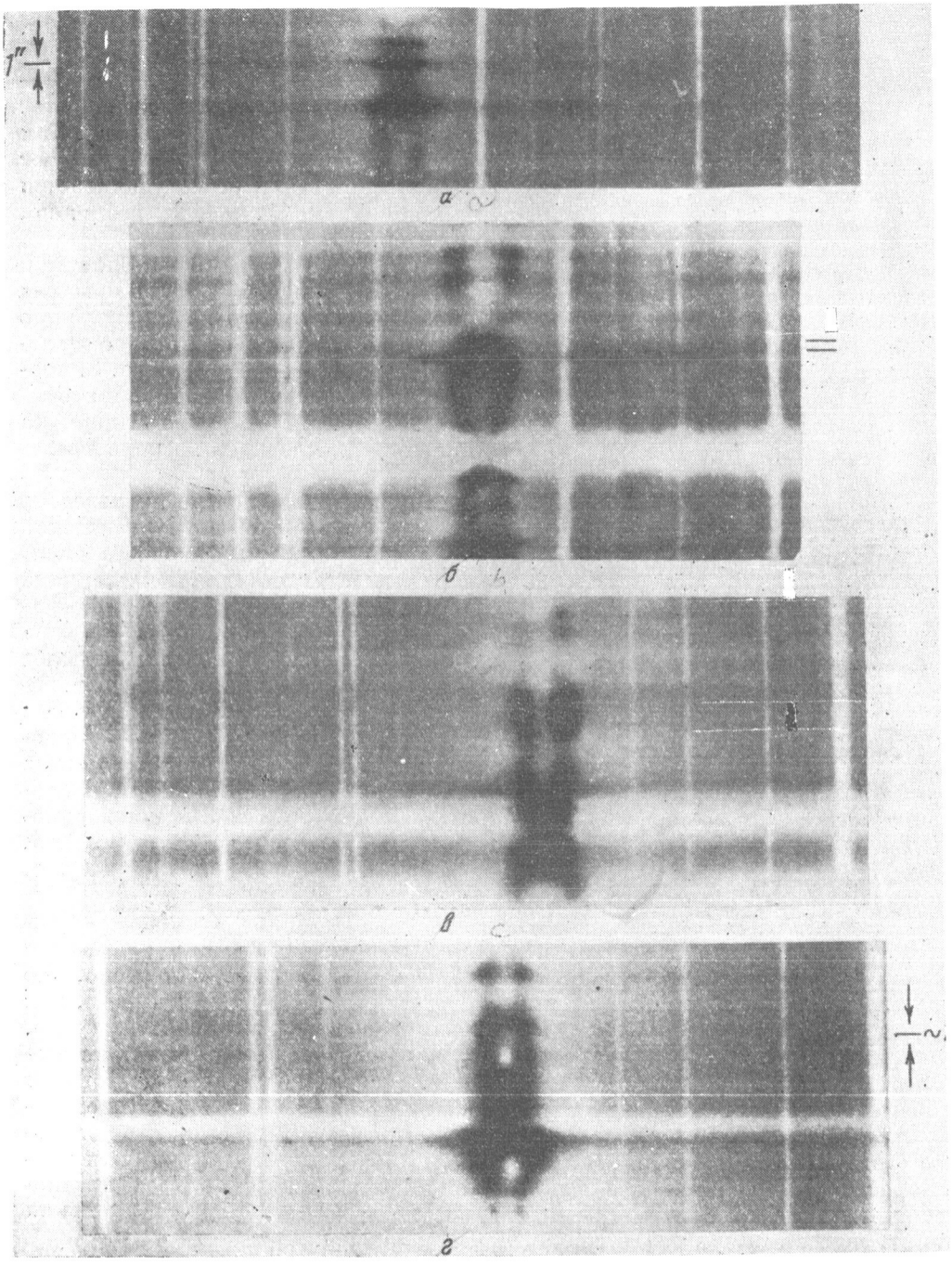


FIGURE 53

The first detailed study of this phenomenon was carried out by Severnyi /10/. At the Crimean Observatory he obtained spectrograms of faculae in various parts of the H_α and H and K lines under conditions of a quiescent image. Severnyi observed that near faculae "Ellerman-bombs" were very frequent. The emission was concentrated in narrow bands of gradually decreasing intensity, which were sometimes almost detached from the core of the line. This showed up most distinctly in the H and K lines; these bands were called "whiskers." (See Figure 53: a) shows the whiskers (top) and the flare emission (bottom) in the H and H_ε lines; b) shows the fine structure of the flare emission in the H and H_ε lines, the emission of the wings consisting of two extremely thin filaments 0".5 wide; c) shows the one-sided "blue whiskers" in the K line near the solar limb; and d) shows the whiskers near the sunspot in the K line.) "Whiskers" are sometimes observed in absorption rather than in emission. The appearance of "whiskers" indicates the existence of gas jets which extend over areas measuring from 700 to 1500 km (from 1 to 2" on the solar disk), and having a wide spectrum of velocities, from values close to zero up to values of hundreds of kilometers per second.

The dimensions specified, which characterize the width of the "whiskers," are apparently exaggerated due to the effects of solar-image vibrations. The actual widths of the "whiskers," according to Severnyi, do not exceed 300 km.

It is interesting that for quiescent solar images ordinary chromospheric flares as a rule split up into distinct small emission centers and "whiskers." The "whiskers" are sometimes one-sided, showing that the gas flows out in one direction only. Mostly, however, they occur in both wings of the line, showing gas efflux in opposite directions. This led Severnyi to attribute the formation of "whiskers" to some type of explosive process, although this would seem to contradict the wide velocity spectrum in the gas. In spite of the brevity of the process, it nevertheless appears to be continuous over a period of a few minutes.

A particularly important factor which may throw some light on the nature of this phenomenon is the formation of narrow emission bands in the continuum. The widths of these bands are the same as those of the "whiskers." One such band is clearly visible in the bottom photograph of Figure 53. The emission of these bands is not temperature emission. The equivalent widths of the Fraunhofer lines intersected by a band are always reduced; the lines appear to be blurred by the continuum emission, regardless of whether they pertain to ionized or to neutral metals. This emission may originate during the braking of fast, "relativistic" electrons in a magnetic field (synchrotronic radiation) or during the braking of fast electrons in an ion field (brehmsstrahlung).

Severnyi's measurements and calculations show that the continuum emission originates in an optically thin layer. This emission appears next to the "whiskers" but almost never represents a direct continuation of the "whiskers." The two phenomena constitute different stages of a single process (or of different, but interrelated, processes) observed both in flares and in faculae. Therefore, we can consider flares and faculae as related phenomena. Moreover, flares are no longer regarded as exceptional phenomena and they are known to be typical processes which occur almost every minute (at least during years of maximum solar activity) on the surface of

the sun. In addition, this explains the appearance of a bright continuum emission in some flares, such as was first observed by Carrington in 1859.

An interesting attempt to interpret the "whiskers" as characteristic "polar auroras" on the sun, produced by corpuscular streams, was made by Shklovskii. He observed that, due to the high velocities in the "whiskers" and the small dimensions within which the motion occurs, gas "blobs" cannot emit in the hydrogen lines for more than 1 sec. Once they have emerged from the "whisker" region, the "blobs" no longer emit. This means that the supply of luminous "blobs" inside the emission zone is continually replenished every few minutes.

The upper levels of the moving hydrogen atoms are excited due to collisions with the "stationary" atoms of the photosphere (that is, with atoms having much lower velocities). Once they have traversed in the chromosphere a distance about equal to the diameter of the "whisker" region, the "blobs" decelerate and finally dissipate. The very existence of the moving gas "blobs" is attributed by Shklovskii to modifications of the magnetic fields near sunspots. Such modifications may release a considerable amount of magnetic energy, able to impart quite high velocities to individual elements of the photosphere (or chromosphere).

§41. Geoactivity and the nature of flares. In order to explain the nature of the processes manifested by chromospheric flares, it is very important to consider the influence exerted by flares on the terrestrial atmosphere and the geomagnetic field.

The first known effect demonstrating the influence which flares have on the ionosphere was the familiar Dellinger effect, which consists in a variation of radio-reception conditions (usually a short-period interruption of the pickup) at the commencement of a flare. Further studies have shown that the composition of the ionosphere, the ionization of the atoms, and even the structure of the ionized-gas layers, all change during flares. These changes are confined almost exclusively to the D layer. The reason for the increased ionization in the D layer may be just the hard ultraviolet and X-ray emission of the flares; hard ultraviolet radiation is emitted by a flare in the form of Lyman quanta, especially L_α quanta.

The main factors are the X-rays originating in the corona above the flares. Rocket observations in the USA and the USSR have shown that the increase in X-rays corresponds exactly to the disturbances of the ionospheric D layer. During chromospheric flares the X-ray flux in the 3 to 8 Å interval at the boundary of the earth's atmosphere increases from 10^{-6} erg/cm^2 · sec to 10^{-2} erg/cm^2 · sec. This radiation-flux increment corresponds to a coronal temperature rise of about 10^7 deg above the flares. Comparable and even greater coronal heating in conjunction with flares is indicated by radio observations (see Chapter IX). By varying the state of the ionosphere as a conducting layer, the hard radiation of flares simultaneously causes slight variations in the geomagnetic field; these take the form of crochets, or small peaks.

The corpuscular radiation of a flare consists of a flux of particles with average velocity of about 1500 km/sec and a flux of relativistic-velocity particles with energies of about 10^9 eV. Strictly speaking, the first type of particles are called corpuscules, while the second type are generally referred to as cosmic-ray particles.

Solar corpuscules cause magnetic storms, polar auroras, and, to a certain extent, all other phenomena in the earth's atmosphere which are

related to solar activity. The presence of these corpuscules (mostly protons) in the earth's atmosphere during polar auroras was inferred from the appearance in the auroral spectrum of Balmer lines which showed a strong violet shift /11/. The shifted line appears as a band which is blurred on the short-wavelength side, indicating that the corpuscules hitting the earth's atmosphere have a continuous velocity spectrum from 1000 to 3000 km/sec. The particular nature of the fluctuations of the geomagnetic field indicates that the corpuscular radiation does not constitute a continuous flux; rather it propagates in the form of comparatively small clouds of corpuscules. It is likely that these clouds originate in the "whiskers."

Severnyi observed directly the motion of a stream of corpuscules through the surge of a flare on 10 May 1950. The surge moved with the velocity of the sources of the solar radio emission. The emission maximum in the surge shifted upward with a velocity of about 800 km/sec. Obviously, this maximum was produced by the corpuscules passing through the surge, whose average velocity (although it was greater than the escape velocity for the sun (562 km/sec) at a distance of 1.1 $R_\odot$ from the photosphere) was somewhat lower than the ordinary corpuscular velocity.

The ejection of corpuscules is apparently also related to strong bursts of solar radio emission, during which radio sources move upward from the photosphere with velocities of hundreds of kilometers per second. Unfortunately, the velocity estimates for such cases are purely qualitative, since they are based on the primitive assumption that radio emission of a given frequency is generated at a definite height above the photosphere.

Cosmic-ray particles are ejected in considerable quantities only during certain large flares. Figure 54 illustrates this point very clearly, for the flares of 28 February and 7 March 1942. These flares caused a radio fadeout, which according to the records of the Godhavn, Cheltenham, and Christchurch stations was followed after 1.5 or 2 hours by a sharp rise in the intensity of the cosmic radiation. A magnetic storm caused by the ordinary corpuscules commenced 20 hours after the flare of 28 February. In other cases the time lag of the cosmic-ray increase relative to the flare commencement was also 1 or 2 hours. Such a large time lag shows that the cosmic-ray particles are ejected (released) from the sun not at the commencement of the flare but during the decreasing part of its development. Nevertheless, they may actually form at the moment of flare maximum or else in the initial stage of the flare development. The intensities of the particle fluxes beyond the earth's atmosphere were about 1 particle per cm^2 per sec. The high neutron-flux intensities observed at stations during flares show that the emission of cosmic rays by the sun continues for a few hours. The neutrons represent secondary particles. The flares of 23 February and 31 August 1956 were particularly strong, and the large 31 August flare was Class 3^+ according to observations at the Crimean Astrophysical Observatory. The flare spectrum included the emission lines of some metals, showed emission reversal of the D_1 and D_2 lines of sodium, and contained the D_3 absorption line of helium. However, only the flare of 23 February produced a cosmic-ray increase. The propagation of the cosmic particles is apparently highly anisotropic, so that they by no means always can reach the earth.

The construction of a theory of flares involves considerable difficulty. The possibility that flares may be electric discharges in the chromosphere

was considered by Alfvén, Giovanelli, Hoyle, and others /12/. In addition, Severnyi /10/ has recently explained flares as a compression of the magnetic field. All flares originate near sunspots, in regions with considerable magnetic fields. Moreover, observations of the flare of 31 August 1956 at the Crimean Astrophysical Observatory showed that a magnetic field of about 500 gauss existed in the flare proper.

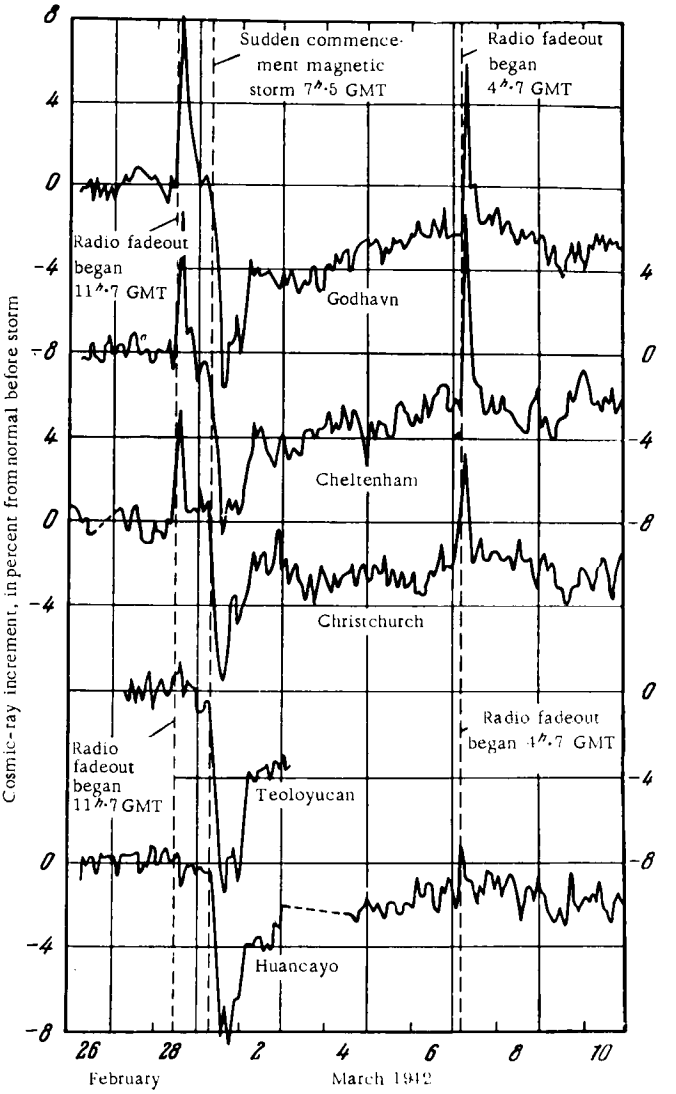


FIGURE 54

Recently Bumba carried out a comparison between 69 flares and the magnetic fields in the active regions where they originated. He came to

the following conclusions:

1) in the first phase of their development flares lie approximately at the polarity boundaries of the group magnetic field;

2) the main flare center generally lies near sunspots consisting of umbrae of different polarity or several (sometimes two) close spots of opposite polarity;

3) the motion of gas in flares develops mainly along neutral field lines, with the field restricting the motion in the flares;

4) a second flare center often forms in a "bay" of the largest spot in the group;

5) about 30% of the flares studied were observed either simultaneously in two neighboring groups or else first in one group and then in another within an interval of a few minutes.

All this points to a relation between the flares and the local magnetic fields in active regions. These conclusions, however, are not entirely indisputable. Many flares form not on neutral field lines but rather in regions of high H or in regions of a high gradient of H. After analyzing 10 large flares accompanied by cosmic-ray emission, Ellison, McKenna, and Reid /13/ concluded that most of these originated above sunspots outside the neutral zones. On 5 and 6 April 1960 at Pulkovo, Krat and Sobolev observed an intense flare which flared up several times in succession in the umbra and penumbra of a sunspot, practically in the upper part of the solar photosphere.

Cowling /12/ has shown that theoretical considerations used to justify the interpretation of flares as discharge phenomena are inconsistent.

Severnyi /14/ suggests that a flare is a manifestation of the pinch effect, a self-compression of the electromagnetic plasma occurring when instability sets in. In this case, the magnetic energy of the plasma is converted into kinetic energy. The very rapid compression, which raises the gas temperature up to 10^7 degrees, generates shock waves propagating from the pinch. These shock waves, which then heat the chromospheric gas to a certain extent, are observed as a chromospheric flare.

At any rate, it must be admitted that the appearance of a chromospheric flare involves an actual explosion. The construction of a theory describing such explosions, however, remains a task for the future.

REFERENCES FOR CHAPTER VI

1. ELLISON, M. A. — M. N., 109: 3. 1949.
2. WALDMEIER, M. — Zs. Aph. 16: 278, 1938; Astr. Mitt., Zurich, No. 153, 1948.
3. DODSON, H. W. and R. R. McMATH. — Ap. J., 115: 78. 1952.
4. KRAT, V. A. and V. M. SOBOLEV. — Izvestiya Glavnoi astronomicheskoi observatorii, No. 162. 1958.
5. RICHARDSON, R. S. and R. MINKOWSKI. — Ap. J. Vol. 89: 347. 1939.
6. ALLEN, C. W. — M. N., Vol. 100: 536. 1940.
7. SEVERNYI, A. B., N. V. STESHENKO, and V. I. KHOKHLOVA. — Astronomicheskii zhurnal, 37: 23. 1960.
8. MUSTEL', E. R. and A. B. SEVERNYI. — Izvestiya Krymskoi astrofizicheskoi observatorii, 5: 3. 1950; 8: 19. 1952; 8: 82. 1955.
9. KRAT, V. A. and L. M. PRAVDYUK. — Izvestiya Glavnoi astronomicheskoi observatorii, No. 159. 1958.
10. SEVERNYI, A. B. — Izvestiya Krymskoi astrofizicheskoi observatorii, 17: 129. 1957.
11. MEINEL, A. B. — Ap. J. 111: 555. 1950; 113: 50. 1951; MEINEL, A. B. and C. Y. FAN. — Ap. J., 115: 330. 1952; 118: 205. 1953.
12. The Solar System, Edited by G. Kuiper, Vol. I. The Sun, Chapter 8. — Izdatel'stvo inostrannoi literatury. 1957. [Russian Translation from English.]
13. ELLISON, M. A., S. M. McKENNA, and J. H. REID. — Publ. Dunsink Obs. I. No. 3. 1961.
14. SEVERNYI, A. B. — Izvestiya Krymskoi astrofizicheskoi observatorii, 20: 22. 1958; 22: 12. 1960; Astronomicheskii zhurnal, 35: 335. 1958; 36: 972. 1959; Doklady Akademii nauk SSSR, 121: 819. 1958.

V. A. KRAT

N67-15491

Chapter VII
THE SOLAR CORONA

Although there are still many "blank spots" in our picture of the solar corona, the advances made in solar physics during the last 15 years have enabled us to establish the main properties of coronal processes. During this time the coronal spectrum has been interpreted and the nature of the transient variations in the corona has been studied. In addition, our ideas of the coronal structure have undergone considerable evolution. From a rather naive model of the corona as a homogeneous gaseous medium in a state of hydrostatic equilibrium, we have proceeded to a more sophisticated concept of a physically inhomogeneous corona in which nonstationary, dynamic phenomena occur.

Fifteen years ago, observations made during total solar eclipses still represented the principal source of information on the corona. Now, however, due to the rapid development of extra-eclipse observation techniques, total solar eclipses no longer are the only effective means of making progress in solar physics, and it is clear that systematic observations using extra-eclipse coronagraphs will give an incomparably greater amount of material than will sporadic studies of the corona during eclipses. Nevertheless, eclipse observations still give valuable information on the low-emission regions of the outer corona and on the weak lines of the coronal spectrum.

The corona is generally divided into two zones, inner and outer. The main factor governing such a division is the presence of emission lines in the spectrum of the inner corona and the absence of these lines in the outer-corona spectrum. Since the coronal emission lines gradually weaken with the distance from the photosphere, there is naturally no sharp boundary between inner and outer corona. Photographs of the corona taken during total eclipses do not actually enable us to make divisions of any kind, since the structural features of the corona (rays, etc.) originating in the inner corona extend into the outer corona without undergoing any discontinuities or even any appreciable changes. There is just a gradual drop in brightness and a smooth change in the geometrical shapes of the structural features as the distance from the disk limb increases.

In addition to this "structural" corona, a diffuse background also exists, which possesses spherical symmetry relative to the center of the sun. The brightness of this background, which (as will be shown in the following) is formed when sunlight is scattered by interplanetary dust grains, drops with an increase in the distance from the sun's disk.

§42. The continuous spectrum of the corona. In order to understand the nature of the solar corona, we first must study its spectrum. The spectrum of the inner corona is a continuum with only a few weak emission lines

(Figure 55, top). No emission lines are observed in the outer corona, which however does include the ordinary Fraunhofer lines (Figure 55, bottom); these Fraunhofer lines, though, are shallower than those in the spectrum of the reversing layer. Grotrian (1934), in his analysis of the data obtained during the 1923 eclipse, attributed the Fraunhofer lines to the diffuse background of the corona, the light of which is unpolarized. This diffuse background constitutes the so-called F component of the corona. The emission of the true "structural" corona is called the K component. In the outer corona, the F component is almost ten times as bright as the K component, while nearer to the disk limb the situation is reversed. In the mixed light of the background plus the corona proper, the equivalent width (w) of any Fraunhofer line is given by the ratio

$$w = \frac{w_0 I_F}{I_F + I_K}, \qquad (1)$$

where w_0 is the equivalent width of the line in integrated sunlight, I_F is the continuum intensity of the F component near the given line, and I_K is the intensity of the K component.

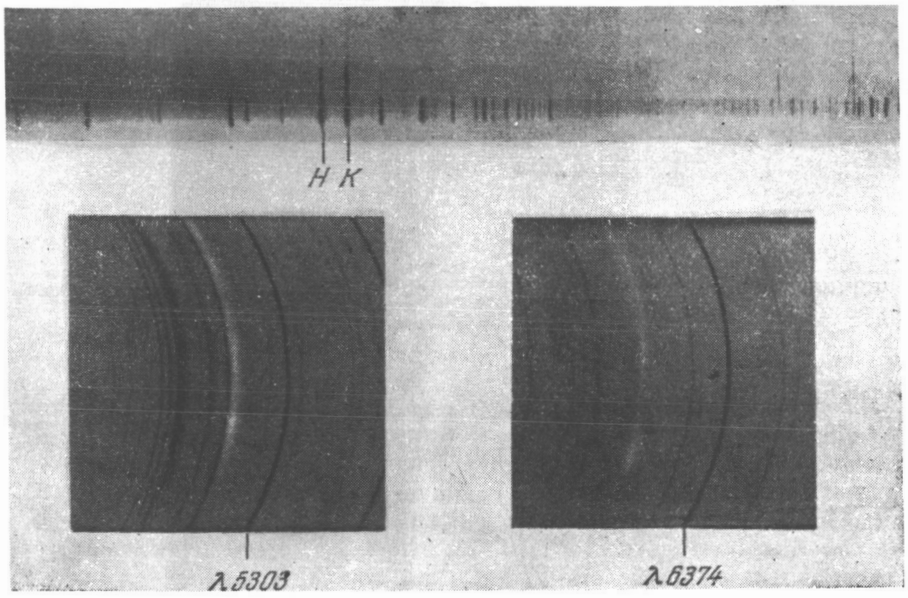


FIGURE 55

If $I_F + I_K$ is known, then equation (1) makes it possible to determine I_F and I_K separately for a given line width w. Near the limb of the solar disk, traces of the deepest Fraunhofer lines are observed only in the long-wave region of the spectrum, especially traces of the D_1 and D_2 sodium lines /1/. This may indicate that the color of the F component is somewhat redder. Van de Hulst /2/, who explained the existence of the F component as being due to light diffraction by interplanetary dust grains located between the sun and the earth (the "interplanetary" solar aureole), showed that the energy

distribution in the F spectrum should be almost identical to that in the integrated Fraunhofer spectrum of the sun. Moreover, Tikhov /3/, who repeatedly studied the color of the solar corona during eclipses using simultaneous photographs taken through different light filters, finally concluded that there is no appreciable difference between the color of the corona and that of integrated sunlight. Tikhov's results apply mainly to the outer corona.

The K component comprises the integrated sunlight scattered by the free electrons of the corona. This explains the partial polarization of the K component. When scattered by electrons, light is totally polarized if the angle of reflection (scattering) is 90°. For reflection angles of 0 or 180°, on the other hand, there is no polarization. Finally, for intermediate cases the polarization is partial. We will return to this question in §43.

Scattering by free electrons should reproduce the solar Fraunhofer spectrum. Nevertheless, the K component does not include any Fraunhofer lines. This is due to the high electron temperature of the corona. The narrow Fraunhofer lines in this case have "Doppler" profiles; their intensities i satisfy the relation

$$i = i_0 e^{-\left(\frac{\Delta\lambda}{\Delta\lambda_D}\right)^2},$$

where $\Delta\lambda_D = \frac{\lambda\xi}{c}$, λ is the wavelength, c is the velocity of light, and ξ is the mean-square thermal velocity of the electrons, defined as

$$\xi = \sqrt{\frac{2kT}{m}} = 5.5\,T^{\frac{1}{2}}\ \text{km/sec}.$$

Even for T values of the order of 10^4, the quantity $\Delta\lambda_D$ for lines with $\lambda \approx 4000$ Å is close to 7 Å. With such a large amount of broadening, the narrow lines will merge into continuous "blends." As far as the wide lines are concerned (for example, H and K of ionized calcium), the calculations of Shain and subsequently of Shklovskii have shown that these lines will be entirely "washed out" against the common background if T is not less than 10^6 degrees. This conclusion, however, exaggerates the situation somewhat. In reality, the H and K lines, before actually becoming so shallow that their central intensities will be indistinguishable from the continuous background within the margin of observational error, will merge at their wings with the "blends" of the many narrow lines, so that the common background will be lower than the background of the continuous spectrum. Even at T = 600,000°, therefore, it will be impossible to observe the H and K absorption lines.

Wide "blends" of Fraunhofer lines over regions of 100 or 200 Å will experience no further "washing out" and can in fact be observed on coronal spectrograms obtained during total solar eclipses. The energy distribution in the coronal "continuum" should be the same as the so-called averaged solar spectrum, in which the intensities have been averaged with a monochromator over intervals about 100 Å wide. Such a curve was obtained by Pettit (see Chapter I).

The very first spectrophotometric measurements (Grotrian, 1929 eclipse) showed traces of a depression of the coronal spectrum in the ultraviolet region, the lowest point of the depression being near 4000 Å. Actually, the intensity of the blends should increase somewhat as the wavelength drops from

4000 Å into the ultraviolet, reaching a maximum near 3800 Å. In accordance with this, Allen (1940 eclipse) observed a depression in the coronal spectrum near 3750 Å. Moreover, he pointed out a depression in the 4030-4200 Å interval, which however is difficult to explain. According to the observations of V. A. Krat (1945 eclipse), an appreciable depression exists only in the 3600-3800 Å interval.

In general, due to the crowding of Fraunhofer lines in the ultraviolet, the K component shows up as a continuous spectrum of a source with a temperature somewhat lower than the spectrophotometric temperature of the sun.

§43. The line spectrum of the corona. Quite complete data on the coronal line spectrum over a wide range of wavelengths can so far be obtained during total solar eclipses only. Since 1933 the coronal spectrograms obtained have been calibrated photometrically, making it possible to determine the line intensities. Until recently, these intensities were determined using a relative scale of so-called "equivalent widths." However, the "equivalent widths" (w) of coronal lines do not correspond exactly to the ordinary concept of an equivalent width. Rather, they are defined as the total intensities of the emission lines divided by the continuum intensity of the corona:

$$w = \frac{\int I \, d\lambda}{I_{c.s.} \cdot 1\,\text{Å}},$$

where λ in the integrand is expressed in Å, and $I_{c.s.}$ is the intensity of the continuous spectrum. Since in general the energy distribution in the continuous spectrum is known satisfactorily, therefore given w we can calculate to within some constant factor the total line intensities in energy units. An absolute standardization was carried out in the observations of the coronal spectrum made during the 1945 eclipse; here, the total line intensities were referred to the continuous spectrum of the solar-disk center.

The results of various determinations of the relative intensities of coronal lines during eclipses, in addition to some intensity estimates obtained using an extra-eclipse coronagraph, are listed in Table 1.

We see that the relative intensities according to different authors vary between wide limits. According to Lyot's extra-eclipse observations, the intensity ratios of the lines $\frac{\lambda 5303}{\lambda 6374}$ vary over two orders of magnitude [102], often falling below unity. This also applies to the ratios $\frac{\lambda 4086}{\lambda 3987}$ and $\frac{\lambda 5694}{\lambda 5303}$ (the latter ratio varies particularly sharply). The λ 5694 line is usually not observed in the coronal spectrum, whereas in some places it may even be stronger than λ 5303 (Waldmeier). The large variability of the coronal lines renders futile any attempt to compile reference tables showing the average line intensities.

The variability of the coronal lines is easy to explain, provided we identify these lines with the lines of multiply ionized atoms. Such an identification was made by Edlén in 1942. Edlén's identification, however, is neither complete nor conclusive for many coronal lines. The strongest lines have been identified correctly, but this is by no means the case for the weak lines. In some instances, the lines of different ions may have quite close wavelengths. In 1950 Shklovskii questioned whether the λ 5694 and λ 5444 lines should really be identified with Ca XV. He even succeeded

in attributing these lines to Ne VIII. Later, however, it was found that the original identification of Edlén was correct after all.

TABLE 1

λ	1	2	3	4	5	6	7	8
3329.6	III	Ca XII	589	0.7			0	
3388.0	II	Fe XIII	325	11.1			26.6	
3454.3	I			1.4			13.9	
3534.0	I?							
3600.9	III	Ni XVI	455	1.1			2.5	
3642.8	III?	Ni XIII?	350?				3.5	
3800.7	I				0.7	0.2	2.5	
3987.3	I	Fe XI	261	0.2	1.1	0.2	2.4	
4086.5	IV	Ca XIII	655	0.3	0.2	0.1	0.5	
4232.2	II	Ca XII	318	0.8	0.7	0.5	1.65	
4311.4	II	Ca XV?					0 9	
4358.9	IV	A XIV?	682				1.6	
4412.4	IV	A XIV?	682					
4566.6	I			0.3	0.3		0.2	
4586	III							
5116.1	II	Ni XIII	350	1.2		0.5	0.6	2.6
5302.9	II	Fe XIV	355	27.5	18.3	(10)	21.0	12.0
5444.5	IV	Ca XV	814					
5536	III	A X	421					1.5
5694.5	IV	Ca XV	814					28
6374.5	I	Fe X	233	2.8	9.4	3	17.7	3.3
6701.9	III	Ni XV	422	2.0	1.8			4.0
7059 6	III	Fe XV	390					29
7891.9	I	Fe XI	261					1 3
8024.2	III	Ni XV	422					240
10746.8	II	Fe XIII	325					150
10797.9	II	Fe XIII	325					

Note. 1) group; 2) ion; 3) ionization potential; 4) equivalent line widths according to Grotrian (1929 eclipse); 5) the same, according to Shain (1936 eclipse); 6) the same, according to Allen (1940 eclipse); 7) equivalent line widths in 10^{-5} Å for continuous spectrum of solar-disk center, according to V. A. Krat (1945 eclipse); 8) equivalent line widths in 10^{-6} Å for continuous spectrum of solar-disk center, according to Lyot's extra-eclipse observations.

Lyot (1939) was the first to attempt some classification of the coronal lines, and he divided the lines into three groups. The number of these characteristic groups has now been increased to four.

The lines of each group behave more or less the same in the spectrum. They are all strengthened or attenuated almost simultaneously. Column 1 of Table 1 indicates the group to which each line belongs. In general, the lines in a single group belong to ions with close ionization potentials. Only for groups III and IV can the ions differ appreciably in ionization potential. Strictly speaking, group IV could be divided up into two groups, but since the number of lines is so few this is not practicable. The intensity ratios of lines within a single group, although they are not always constant, still fluctuate less than do the intensity ratios of lines from different groups.

Shain (1936 eclipse) and then Allen (1940 eclipse) established that on the whole the intensity of the coronal lines drops more sharply with the distance from the solar limb than does the continuum intensity. Starting from a distance (ϱ) of 2 disk radii from the disk center, which is generally taken as unit distance, no coronal lines were observed. Allen (1940) made photometric measurements of the green coronal line which extended farther out than the others, reaching to distances of $\varrho = 1.8$ from the limb.

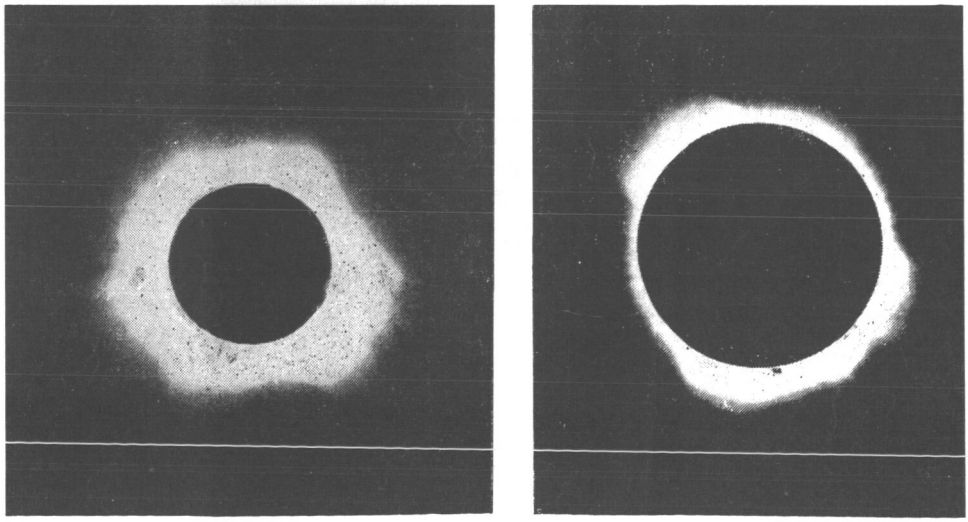
§44. **The brightness and polarization of the corona.** The brightness of the K component fluctuates greatly, due to the complex structure of the corona. Nevertheless, in order to form a general picture of the brightness attenuation with the distance away from the solar limb, Baumbach used the results of observations during total solar eclipses to derive a certain correlation between the average brightness B (averaged over the data of various years) and ϱ. The quantity B as a function of ϱ can be expressed by the simple formula

$$B(\varrho)=\frac{0.0532}{\varrho^{2.5}}+\frac{1.425}{\varrho^{7}}+\frac{2.565}{\varrho^{17}}, \qquad (2)$$

where the brightness at the center of the solar disk is taken as 10^6. Expression (2) does not allow for the variations in the corona during the solar-activity cycle. The gradient of B with ϱ is in general quite variable (Mikhailov, Bečvář, and others). Formula (2) gives the total brightness of both the K and F components; Bogorodskii and Khinkulova /4/ eliminated the F component and then used Baumbach's formula to obtain the following interpolation formula for the K component alone:

$$B_K(\varrho)=\frac{0.021}{\varrho^{3}}+\frac{1.33}{\varrho^{7}}+\frac{2315}{\varrho^{17}}+\frac{0.349}{\varrho^{40}}. \qquad (3)$$

Absolute photoelectric measurements of the illumination from the zone within the limits $1.030 < \varrho < 6.00$, carried out during eclipses by Kuntz and Stebbins (1918 and 1925), Briggs (1922), Stebbins and Whitford (1937), and Nikonov and Nikonova (1945), show that during years of maximum activity the luminous flux from the corona is approximately twice that during epochs of minimum.



a b

FIGURE 56

In 1887 Ganskii showed that during years of maximum activity the corona is comparatively rounded, whereas during years of minimum the corona becomes definitely "elongated" in its equatorial plane (Figure 56, a, b).

By applying the law describing light scattering by free electrons, we can turn the function $B_K(\varrho)$ in (3) into a function expressing the dependence of some averaged concentration of free electrons (n_e) upon ϱ. The coefficient of Thomson scattering per electron is given by the familiar formula

$$\sigma(\psi) = \frac{1}{2}\frac{e^4}{m^2c^4}(1+\cos^2\psi), \tag{4}$$

where ψ is the scattering angle. Taking into account the variation of σ with ψ, Bogorodskii and Khinkulova calculated the function $n_e(\varrho)$ from $B_K(\varrho)$. The function $n_e(\varrho)$ can be represented by the simple binomial formula

$$n_e(\varrho) = 10^8\left(\frac{2.13}{\varrho^6}+\frac{3.05}{\varrho^{25}}\right). \tag{5}$$

The values of $n_e(\varrho)$ were calculated for ϱ from 1.0 to 2.6.

However, the reader should be cautioned against an uncritical application of formula (5) or of tables expressing this dependence. The structural inhomogeneities of the corona and the variations of n_e with the phase of the solar-activity cycle are so great that the true n_e for a given ϱ may differ by a factor of ten or more from the figures calculated according to (5). Van de Hulst's attempt to obtain a better average function for $n_e(\varrho)$, by taking $n_e(\varrho)$ into account separately for the equatorial and polar zones of the corona and by distinguishing between the data obtained during epochs of maximum and minimum solar activity, can by no means be regarded as successful, since the initial values of $B_K(\varrho)$ for different epochs differ to such an extent that typical cases can be picked out only quite arbitrarily. Qualitatively, van de Hulst's calculations do not differ from those of Bogorodskii and Khinkulova. The most valuable part of van de Hulst's calculations is his estimate of the variation of the degree of polarization of the corona with ϱ. The average degree of polarization is, unlike $n_e(\varrho)$, of considerable worth to observers, since it helps them estimate the true continuum intensity and the accuracy of such estimates, in cases when observations of the "white" corona (the polarized component of the continuous emission of the corona) are made under extra-eclipse conditions using polarizing photometers (Dollfus, Wlérick).

The polarization at some particular point of the coronal image depends on the angular distribution of the scattered light. In the inner corona, it also depends on the law of limb darkening, since due to this phenomenon light beams from different zones of the photosphere reach the corona with different radiation densities. The structure of the corona has a strong influence on the polarization, and the degree of polarization therefore depends on the spatial orientation of the coronal formations relative to the observer's line of sight. This factor makes it unnecessary, and practically impossible, to obtain an exact theoretical prediction of the degree of polarization at any desired point on the coronal image. The theoretical calculation of the degree of polarization for different epochs of solar activity (equally for the maxima and minima) carried out by van de Hulst for the equatorial zone of the corona gives a degree of polarization (p) of 17% near the solar limb, with values of 40% for $\varrho = 1.5$ and 6 to 10% for $\varrho = 4.0$. The determinations of ϱ during solar eclipses are on the whole consistent with the theory (deviations generally do not exceed 10%).

§45. Structural features of the corona in "white" light. Since the corona is a combination of unstable and inhomogeneous (relative to T and n_e) formations, §42 and §43 can serve just as a preliminary introduction to the problem. Actually, coronal research gets under way only with an analysis of the structural features of the corona.

Historically, the corona was first investigated using photographs taken in "white" light during total solar eclipses. The adjective "white" should here be interpreted in the sense that the corona was photographed not in the wavelengths corresponding to specific lines, but rather over wide spectral regions (mainly, the photographic and photovisual regions). In this case the radiation flux is almost entirely due to the continuous spectrum, and the brightness distribution of the visible formations is determined by the free-electron distribution (n_e) there. Such ordinary photographs of the corona may be called photographs of the "electron" corona.

The most comprehensive description of the coronal forms is that given by Bugoslavskaya /4/. She specified the following four principal types of formations; 1) fans (or helmets), 2) systems of arches and shells above sunspots, 3) systems of arches above prominences, and 4) rays.

The coronal forms are determined by photospheric phenomena (sunspots, faculae), chromospheric phenomena (spicules, flares), and prominences. The fans (or helmets) are conical formations which taper upward. They form over quiet dense prominences, and the prominence shows up in the central, darker zone of the helmet. Thus, the corona, without actually adhering to the prominence, forms a comparatively thin envelope around it. This envelope is often comprised of several layers (elliptic "arches"), and it extends to heights of $\varrho = 1.5$ or 1.8. Ascending motions with average velocities of about 2 km/sec are observed in it. The envelope is surrounded by a system of "fan" rays, these being thin, curved streamers which, converging to the top of the fan, produce the helmet-shaped contour. The fan axis is generally inclined to the perpendicular to the solar surface. During years of minimum activity, the fan axes are deflected toward the equator at angles of around 30 to 50° to the normal, thus forming the characteristic coronal "wings." The fans are responsible for the variations in coronal shape during the course of the solar-activity cycle, as was first proved reliably by Ganskii (1887). Due to their concentration and deflection toward the equator, the "minimal" corona, contrary to the more nearly spherical "maximal" corona, is greatly elongated in the plane of the equator.

Around spot groups, there are generally "arches" which almost converge at the base. These "arches," just like the fan "arches," actually constitute characteristic shells or envelopes.

The arches over small prominences are apparently similar in structure to the fans, although they are much lower than the latter. These arches are shaped like semiellipses with their centers near the solar limb.

The coronal rays are the most widespread formations. The short, slightly curved rays in the polar regions are, according to Bugoslavskaya, Zel'tser, and Markov (1941 eclipse), closely related to large chromospheric spicules, which point in the same direction as the rays; the coronal rays here appear to be a continuation of the spicules. Rays of this kind, according to Bugoslavskaya, may also be observed above normal, "undisturbed" parts of the photosphere.

Wide beams of rectilinear rays are observed above faculae, and the widths of these beams may reach $0.03 R_\odot$ or more. Near sunspots these

rays are deformed and the beams become "bunched" together. They often have velocities present in them of around 10 km/sec. The rays over faculae, like fans, are very extended formations. They may reach out to beyond distances of several solar radii from the limb.

The shape of the polar rays and their position near the poles (Figure 56 b) are possible indications of the existence of some general magnetic field of the sun. The fans, the arches, and the ray systems can apparently be attributed to local magnetic fields.

A certain similarity between the orientation of the polar rays and that of the force lines of a dipole was at one time even regarded as proof of the dipole nature of the sun's magnetic field. A more detailed study of these rays, however, led van de Hulst /2/ to conclude that they deviate greatly from a dipole pattern. This problem was recently considered in very great detail by Csada /5/, who showed that the solar magnetic field is a dipole field at large distances from the sun, whereas near the surface a reversal of magnetic polarity is observed near the equator; the equatorial polarity in each hemisphere is opposite to the polarity of the respective pole. Such a field may be an indication of the presence of large-scale motions in the solar interior (amplifying and sustaining the field), although it is not impossible, too, that the general field of the sun is some statistical average, determined by the preferred orientation of the local magnetic fields due to solar rotation. The theory of the general magnetic field of the sun, however, has still not been developed to any real extent.

§46. The ionization and excitation of atoms in the corona. A theory explaining the ionization of the corona was first constructed by Shklovskii /6/, who pointed out the absence of thermodynamic equilibrium in the corona. The coronal gas is maintained at very high temperatures, and therefore it is ionized mainly by electron impact. Moreover, since this gas is highly rarefied, the only reverse process possible is photorecombination. Ionization equilibrium is described by the expression

$$n_e n_i Z_i(T_e) = n_e n_{i+1} \sum_{m=1}^{\infty} R_{i,m}(T_e), \tag{6}$$

where n_e is the electron concentration, n_i is the concentration of the ions which have lost i electrons, n_{i+1} is the concentration of the ions denuded of $i+1$ electrons, Z_i is the coefficient of ionization from state i to $i+1$ (ionizations from the ground level only are considered), and $\sum_{m=1}^{\infty} R_{i,m}$ is the sum of the photorecombination coefficients for all the levels of the i th state. Since $Z_i(T_e)$ and $\sum_{m=1}^{\infty} R_{i,m}$ differ for different atoms, there is no single ionization equation (analogous to Saha's formula) which applies to all the elements. Equation (6) may be written more conveniently as

$$\frac{n_{i+1}}{n_i} = \frac{Z_i(T_e)}{\sum_{m=1}^{\infty} R_{i,m}(T_e)}. \tag{7}$$

$Z_i(T_e)$ can be calculated only when the total effective cross sections for ionization by electron impact are known. Unfortunately, effective impact-ionization cross sections have not in general been determined either experimentally or theoretically (except for helium). They can be found using either Thomson's classical formula or else the quantum-mechanical

TABLE 2

Ion \ T_e	$5\cdot10^5$	$6\cdot10^5$	$7\cdot10^5$	$8\cdot10^5$	10^6	$1.2\cdot10^6$	$1.4\cdot10^6$
				$Z(T_e)$			
Fe IX	$3.4\cdot10^{-12}$	$5.5\cdot10^{-12}$	$9.4\cdot10^{-12}$	$1.6\cdot10^{-11}$	$3.4\cdot10^{-11}$	$5.5\cdot10^{-11}$	$7.3\cdot10^{-11}$
Fe X	$1.4\cdot10^{-12}$	$3.3\cdot10^{-12}$	$4.8\cdot10^{-12}$	$9.4\cdot10^{-12}$	$1.9\cdot10^{-11}$	$3.2\cdot10^{-11}$	$4.8\cdot10^{-11}$
Fe XI	$6.0\cdot10^{-13}$	$1.5\cdot10^{-12}$	$2.4\cdot10^{-12}$	$4.8\cdot10^{-12}$	$1.2\cdot10^{-11}$	$2.2\cdot10^{-11}$	$3.5\cdot10^{-11}$
Fe XII	$2.1\cdot10^{-13}$	$6.0\cdot10^{-13}$	$1.6\cdot10^{-12}$	$2.4\cdot10^{-12}$	$6.2\cdot10^{-12}$	$1.2\cdot10^{-11}$	$2.2\cdot10^{-11}$
Fe XIII	$8.8\cdot10^{-14}$	$2.9\cdot10^{-13}$	$8.3\cdot10^{-13}$	$1.8\cdot10^{-12}$	$3.6\cdot10^{-12}$	$8.8\cdot10^{-12}$	$1.6\cdot10^{-11}$
Fe XIV	$3.3\cdot10^{-14}$	$1.2\cdot10^{-13}$	$3.7\cdot10^{-13}$	$8.8\cdot10^{-13}$	$2.4\cdot10^{-12}$	$4.7\cdot10^{-12}$	$8.3\cdot10^{-12}$
Fe XV	$5.5\cdot10^{-15}$	$2.7\cdot10^{-14}$	$1.0\cdot10^{-13}$	$2.7\cdot10^{-13}$	$1.1\cdot10^{-12}$	$1.7\cdot10^{-12}$	$5.4\cdot10^{-12}$
Ni XI	$4.7\cdot10^{-13}$	$7.2\cdot10^{-13}$	$1.8\cdot10^{-12}$	$2.6\cdot10^{-12}$	$7.2\cdot10^{-12}$	$1.3\cdot10^{-11}$	$2.0\cdot10^{-11}$
Ni XII	$1.0\cdot10^{-13}$	$3.0\cdot10^{-13}$	$9.4\cdot10^{-13}$	$1.9\cdot10^{-12}$	$3.6\cdot10^{-12}$	$9.3\cdot10^{-12}$	$1.8\cdot10^{-11}$
Ni XIII	$3.7\cdot10^{-14}$	$1.5\cdot10^{-13}$	$3.9\cdot10^{-13}$	$1.0\cdot10^{-12}$	$2.5\cdot10^{-12}$	$4.9\cdot10^{-12}$	$8.3\cdot10^{-12}$
Ni XIV	$1.3\cdot10^{-14}$	$5.5\cdot10^{-14}$	$1.9\cdot10^{-13}$	$4.9\cdot10^{-13}$	$1.4\cdot10^{-12}$	$3.2\cdot10^{-12}$	$5.5\cdot10^{-12}$
Ni XV	$5.5\cdot10^{-15}$	$2.6\cdot10^{-14}$	$1.0\cdot10^{-13}$	$2.7\cdot10^{-13}$	$1.0\cdot10^{-12}$	$1.6\cdot10^{-12}$	$3.4\cdot10^{-12}$
				$\Sigma R(T_e)$			
Fe IX	$3.2\cdot10^{-12}$	$2.9\cdot10^{-12}$	$2.5\cdot10^{-12}$	$2.2\cdot10^{-12}$	$1.9\cdot10^{-12}$	$1.6\cdot10^{-12}$	$1.4\cdot10^{-12}$
Fe X	4.0	3.5	3.2	2.7	2.4	2.0	1.8
Fe XI	4.9	4.3	4.1	3.4	2.9	2.5	2.2
Fe XII	6.0	5.2	4.7	4.3	3.5	3.1	2.8
Fe XIII	7.1	6.2	5.5	5.1	4.3	3.6	3.2
Fe XIV	8.3	7.3	6.5	5.9	5.2	4.5	3.9
Fe XV	9.8	8.7	7.8	7.0	5.9	5.4	4.9
Ni XI	$5.2\cdot10^{-12}$	$4.4\cdot10^{-12}$	$4.0\cdot10^{-12}$	$3.6\cdot10^{-12}$	$3.1\cdot10^{-12}$	$2.6\cdot10^{-12}$	$2.2\cdot10^{-12}$
Ni XII	6.2	5.4	4.8	4.5	3.8	3.2	2.7
Ni XIII	7.4	6.5	5.7	5.3	4.5	3.8	3.2
Ni XIV	8.7	7.6	6.7	6.1	5.2	4.6	4.1
Ni XV	$9.8\cdot10^{-12}$	$8.7\cdot10^{-12}$	7.7	7.0	5.9	5.3	4.8
Ni XVI	$1.4\cdot10^{-11}$	$1.0\cdot10^{-11}$	8.6	8.2	6.8	6.1	5.5

formula derived for fast electrons. Shklovskii calculated Z using Thomson's formula; the errors introduced are not excessive, so that the correct values for $\frac{n_{i+1}}{n_i}$ are preserved. $\sum_{m=1}^{\infty} R_{l,m}$ was calculated using Menzel's formula for hydrogenlike atoms, this being a satisfactory approximation for the coronal ions. Shklovskii's ionization coefficients (Z) and the total recombination coefficients (ΣR) are listed in Table 2.

From this table it is easy to find out which ions prevail at each electron temperature. For example, at $T_e \approx 600{,}000°$ FeX ions will be the most numerous iron ions, while at $T_e \approx 1{,}200{,}000°$ FeXIV will predominate. It is important to note that the variation of $\frac{n_{i+1}}{n_i}$ with T_e is so sharp that errors in the effective ionization cross sections are not significant as far as this dependence is concerned.

The problem of the excitation of coronal ions, that is, the derivation of the ion distribution over the various energy states, can be (and is being) solved independently of the problem of ionization.

The number of excited ions in the line of sight is easy to find using the familiar formula

$$N_k = \frac{4\pi I_\lambda}{A_{kl} h \nu_{kl}}, \qquad (8)$$

where I_λ is the absolute intensity of the coronal line in ergs/cm^2 · sec, A_{kl} is the transition probability from the upper level k to the lower level l, and ν_{kl} is the transition frequency.

The concentrations of atoms in excited states, and thus the numbers of ions in ground states, were calculated by Kun Huang, Shklovskii, and Yakovkin. Electron impacts and spontaneous transitions from upper levels are the factors which increase the population of any given level. Electron impacts of the second kind and spontaneous transitions to lower levels, on the other hand, reduce the population of the level. The radiation excitation for the inner corona, where the coronal emission lines originate, is insignificant. For example, for the λ 5303 and λ 6374 lines excitation by radiation prevails over excitation by electron impact only in the outer corona, starting with $\varrho > 2.1$.

The concentrations of coronal ions are too low to permit any noticeable self-absorption in the coronal lines; for this reason they do not show up as emission lines on the solar disk either.

Calculations of the total ion concentrations enabled Shklovskii to make a reliable estimate of the concentration ratio between metals and hydrogen atoms (protons). Although these calculations are by necessity crude, it is clear that, at any rate, no appreciable difference exists between the chemical composition of the corona and that of the reversing layer.

An interesting conclusion follows from theoretical calculations of the line intensities in the far ultraviolet using formula (8), namely that several strong lines exist in the spectral region $\lambda < 1000$ Å. These lines include FeXV, MgX (ionization potential of 390 V), NeVIII (ionization potential of 261 V), and others. Shklovskii has observed that the lines of MgX (λ 625 and λ 610) and NeVIII (λ 776 and λ 768) should be particularly strong.

§47. Profiles of coronal lines. The structure of the corona in monochromatic emission of spectral lines. The first reliable data indicating

large widths for the coronal lines were obtained by Lyot (1939), who measured the halfwidth of the green line (λ 5303) with an extra-eclipse coronagraph. In 1941 these data were confirmed by Kalinyak, who was the first to use an interferometer during a total solar eclipse to investigate the profile of the green line. However, until very recently the data on coronal line profiles constituted merely the results of scattered measurements, which were useful only in that they established that the atoms in the corona move with high velocities due to the high temperature there. Only lately has extensive information been obtained on the profiles of the λ 6374, λ 5303, and λ 5694 lines. According to Dollfus's observations (1942 to 1952), made with a 3.2 Å/mm dispersion but a fairly wide instrument profile (the slit width converted to wavelength units was 0.3Å), the λ 6374 line had a half-width $\Delta'\lambda$ ranging between 0.86 and 0.91 Å; the quantity $\Delta'\lambda$ showed some tendency to increase away from the solar limb. Taking into account that the correction for the instrument profile may not have been right, van de Hulst /2/ assumed that only thermal velocities were responsible for the Doppler broadening, and thus he obtained a temperature T of about $2 \cdot 10^6$ deg. This value of T, however, is quite unacceptable for those regions of the corona where the red line reaches maximum brightness.

Even stranger results for this line were obtained by Billings, Pecker, and Roberts, using a spectrograph with half the dispersion and resolving power as that used by Dollfus. They found T to be between $2.3 \cdot 10^6$ and $1.5 \cdot 10^7$ deg. The high values of T were obtained for a so-called "active" sun, when active regions are located near the limb. The values of T obtained, in spite of their inaccuracy, indicate that the broadening of the red line is of nonthermal origin. Actually, for $T > 2 \cdot 10^6$ the number of FeX ions is too small for the λ 6374 line to be visible. The factor responsible for the broadening of the coronal lines is apparently a turbulence of the gas analogous to that observed in the chromosphere. This conclusion is supported by the results of the most accurate of all the recent investigations of the profiles of the green coronal line, namely that carried out by Prokof'eva at Pulkovo (1956). Prokof'eva, using a spectrograph with 4 Å/mm dispersion and an instrument-profile halfwidth of 0.22 Å/mm, found $\Delta'\lambda$ values ranging between very wide limits (from 0.55 to 1.10 Å).

This high scattering of $\Delta'\lambda$ shows that in the zones of green-line emission the turbulence, although it is somewhat less than in zones of red-line emission, is nevertheless still important. It is particularly noteworthy that, although the temperature in the red-line zones is around half that in the green-line zones, the red-line halfwidth is on the whole even greater than the green-line halfwidth. According to Prokof'eva, the mean turbulent velocity in the green-line zones is from 9 to 33 km/sec. In zones of red-line emission this velocity should be almost twice as high. A different result could be obtained only if we assumed an isothermal corona, in which case the green and red lines would be emitted in the same regions. Such an assumption, however, contradicts Waldmeier's observations over many years, from which it was first established that these coronal lines originate in different zones. Radio observations (see Chapter VI) also indicate that the value of T in the corona is on the average not greater than 1,000,000°.

The only more or less reliable determination of the yellow-line profile (λ 5694) is that of Billings, who worked with spectrograms obtained at the Climax Station using a dispersion of 2.5 Å/mm. The accuracy of the

average profile obtained, however, is not very high, due to errors connected with the influence of the Fraunhofer spectrum of the solar halo. $\Delta'\lambda$ was found to be 1.32 Å (not corrected for broadening in the instrument). When this value was corrected for instrument effects, it gave a temperature of $4.2 \cdot 10^6$ deg (arbitrarily assuming no turbulence and a proposed identification of the line with the CaXV line). This temperature is in full agreement with the theoretically predicted optimum for the production of CaXV ions. If Billings's results are corroborated by future findings, then this will indicate that the turbulence in the yellow-line zones (λ 5694) is very low indeed. At any rate, for the yellow-line profile to remain unaffected by turbulence, the mean turbulent velocities must not exceed a few kilometers per second.

Waldmeier's observations /8/ showed that, whereas the green-line emission is more concentrated in low latitudes and its maxima are observed above active regions, the red-line emission is distributed more uniformly along the solar limb and it doesn't show any sharp correlation with active regions. Moreover, the two coronal lines differ in behavior during the 11-year cycle. On the average the green line reaches a maximum almost simultaneously with the activity maximum, while the red-line emission maximum lags considerably and also retains a fairly high value in the epoch of minimum activity. According to the theory of coronal ionization, this fact should correspond to a certain overall cooling of the corona during the epoch of minimum activity, which results in the appearance of extensive, quite "cool" regions characterized by helium and ionized-calcium emission lines (compare the observations of Gnevyshev and Gnevysheva at the Mountain Astronomical Station near Kislovodsk).

According to Waldmeier and the astronomers at Pic du Midi (Trellis, Dollfus, and others), the most typical coronal forms observed with polarizing interference filters have been found to be coronal rays and coronal condensations.

Coronal rays are zones of maximum coronal-line emission which remain stable for a few days or more. In these zones the line emission also reaches maximum height (relative to neighboring areas). The detailed investigations of Trellis /9/ show that the rotation of the corona, if it is derived according to coronal rays, is different from the rotation of the reversing layer. At the equator the corona rotates somewhat more slowly, while near the poles its rotation is faster. This differential rotation is described by the formula (see Chapter I)

$$\xi = 14^\circ.16 - 1^\circ.89\,\varphi^2. \qquad (9)$$

Here ξ is the diurnal displacement in longitude, and φ is the latitude in radians. The rays are almost invariably latitudinal. The electron concentration in the rays is quite variable; in the brightest rays it may be ten times greater than in the surrounding coronal matter. The data on the electron concentration, however, are very inaccurate. More reliable estimates will be obtained only when sufficient data are accumulated from extra-eclipse observations of the "white" corona, using polarizing photometers. Such photometers, which make it possible to separate the polarized light of the "white" corona from the celestial background, are now employed at the Climax (Colorado) and Pic du Midi Observatories. It has

been established that all the bright rays in the λ 5303 line appear only over facular areas.

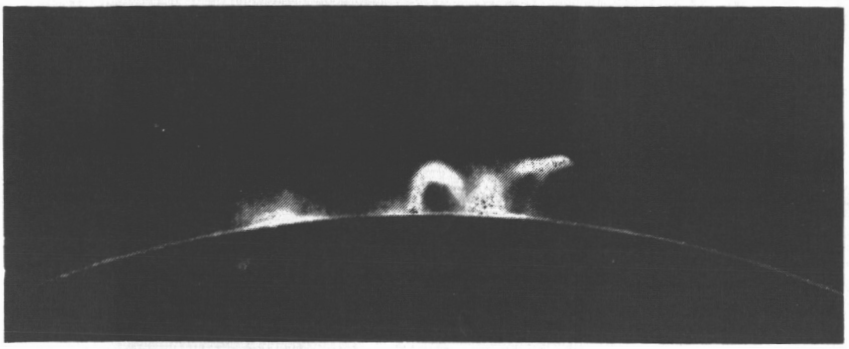


FIGURE 57

Coronal condensations are often domelike in shape, and they show up as arches when seen in cross section (Figure 57). In the green line, individual bright cloudlets are observed near the arches; however, the cloudlets are often diffuse, and they resemble bright clouds or prominences. It has been established (apparently quite reliably) that the yellow line at λ 5694 appears only in coronal condensations. Opinions have varied for a long time with regard to the stability of coronal condensations. For example, Waldmeier, observing that condensations always form over spot groups, suggested that they can exist for a few days. Then, the more recent investigations of Waldmeier, Roberts, and others showed that coronal condensations are closely connected with chromospheric flares. The apparent stability of the condensations thus may be a result of the recurrence of flares in the same region. There is no point in differentiating here between flare condensations and condensations around the so-called sunspot prominences, since flares and sunspot prominences, as well as the so-called loop prominences, just represent different facets of the very same phenomenon. These prominences, when projected onto the solar disk, do not differ from flares (they are much brighter than the background). In cases when no relation can be established between a coronal condensation and some chromospheric phenomenon or other, we can by no means assume that such a relation does not exist. Coronal condensations are observed quite high above the solar limb. The flares corresponding to them may occur on the limb itself, on the boundary with the photosphere, or even somewhat beyond the limb.

According to Lyot and other authors, the coronal condensations may rise to average heights of as much as $\varrho = 1.2$. According to Waldmeier, the enhancement of the continuous spectrum in coronal condensations indicates an increased concentration of free electrons there, up to values of the order of 10^{10}. This continuum enhancement is very slight, and generally does not exceed 10 to 15% of the intensity of the continuous coronal background near the condensation.

Some condensations, particularly those which are diffuse or cloudy in shape, may be related to sunspots. However, such condensations have not yet been studied to any real extent.

§48. Coronal motions. The development of coronal condensations. Given the high eddy velocities in the corona, it is somewhat strange that the macroscopic velocities of the large gas masses there are very low. Dollfus, on the basis of numerous spectrograms obtained at the Pic du Midi Observatory, established that many of the red-line profiles for λ 6374 show no marked Doppler shift. The highest macroscopic velocities in quiet coronal regions did not exceed 2 km/sec. Prokof'eva, in her studies of the green line, obtained somewhat higher velocities, on the average about 5 km/sec. The velocity difference for the red-line and green-line emission zones is quite natural, but in both cases low macroscopic velocities are involved.

Comparatively high velocities which show up as Doppler shifts are observed only near coronal condensations and in the condensations themselves. Waldmeier repeatedly observed coronal arches moving with velocities of about 10 km/sec. According to observations at Climax using a new spectrograph with 2 Å/mm dispersion, Newkirk obtained average velocities of 14 km/sec in these "active" regions. It is clear that whenever large-scale directed motion can be expected in the corona, the corresponding velocities will have the same order of magnitude as the velocities of the prominences.

The absence of large Doppler shifts in the "quiet" zones can be explained only by assuming the presence of extremely disordered small-scale gas currents. An analogy with the chromosphere can be drawn here. In spite of the strong turbulence in the chromosphere, the chromospheric lines formed by the multitude of moving gas filaments do not show any Doppler shift at all. Only in cases of ordered chromospheric motion (such as in flares, in the cold chromospheric condensations recently discovered by T. V. Krat, and in the chromospheric faculae) is there a noticeable Doppler shift in the chromospheric lines as a whole.

Of particular interest is the development of coronal condensations. In the presence of chromospheric flares, the development of a condensation represents a very complex process, since a large flare generally appears to consist of several successive, or even simultaneous flares.

The most detailed and clear description to be had at present is that of the development of condensations over loop prominences. Motion-picture films obtained at the Climax Observatory (Roberts and others) show that the streamers from a loop prominence which appear to pour continually into the photosphere are formed from coronal matter. The appearance of a condensation precedes the development of the prominence somewhat, the yellow-line emission being the most intense at the beginning. When the flare decays, this line disappears, although the emission lines at λ 5303 and λ 6374 remain strong until the end. At first, however, these lines often disappear entirely in the condensation, as can best be diagnosed from the enhancement of the continuous spectrum (Waldmeier). Finally, both the coronal condensation and the flare disappear. The development of a condensation can be conceived of most easily as a compression (possibly due to changes in the magnetic field) of coronal matter, followed by a cooling and an efflux of gas into the photosphere. In other words, the formation

of a coronal condensation represents a transformation of coronal matter into photospheric gas.

§ 49. **The corona and the chromosphere.** Numerous slow-motion films made in H_α light using polarizing interference filters have shown that an intensive exchange of matter takes place between the corona and the chromosphere. Prominences are often seen to originate in the corona and subsequently to fall into the chromosphere. The opposite is also observed, in which case the prominence breaks up and vanishes into the corona. With respect to these interchanges, there are still two problems which have to be solved: 1) the problem of the heating of coronal gas, and 2) the problem of the dissipation [or evaporation] of gas from the corona.

It is much easier to explain why the corona remains very hot than to give a correct quantitative interpretation of the coronal heating. Thus, there is no lack of qualitative speculations on the subject.

Simple calculations of the energy lost by the corona to outer space (Shklovskii and others) show that these losses are small. Actually, the corona loses energy mainly via emission in the spectral lines, and the number of such lines and their intensities are small relative to the chromosphere, the spectrum of which is closely packed with emission lines from the far ultraviolet to the infrared. The heating of the corona has been explained by two different hypotheses. The first hypothesis (V. A. Krat) suggests that the corona becomes heated due to collisions with fast prominences and chromospheric spicules, whose average velocities are around 100 km/sec. When gas streamers (spicules and prominences) decelerate and decay by some turbulent and then gas-kinetic means (gas-kinetic viscosity), their kinetic energy is converted to heat. If the velocity of a spicule is v, then the total energy per atom in this spicule is

$$\frac{m_{\mathrm{H}}}{2}v^2+\frac{3}{2}kT. \qquad (10)$$

Here the coronal gas is assumed to consist only of hydrogen, since this is virtually the case for the outer layers of the sun. As we have noted, when a spicule decays, its kinetic energy is converted to heat. Thus, if the average T of a spicule is taken as 10^4 degrees, then after it stops and decays T will rise to 400,000°. Although this gives a correct order of magnitude for the heating of the coronal gas, such a mechanism can hardly supply the power required, due to the rarity of fast ejections from the chromosphere. The situation is further complicated by the fact that a considerable part of the coronal gas cools, producing prominences which pour back into the chromosphere. In order to explain how the average T of the corona remains so high, we must make the energy influx at least equal to the overall energy losses of the corona. Krat's hypothesis can be modified by assuming that the coronal gases are heated by the turbulence in the corona itself. We have seen that the energy of this turbulence, at least for the "red" regions of the corona, is on the average much higher than the thermal energy; therefore the decay of turbulent streams must inevitably produce local heating of the corona. This, however, still does not solve the problem, since a new question arises; namely, what are the forces responsible for the coronal turbulence?

The other hypothesis (Shklovskii and Pikel'ner) is unfortunately no better. According to the second hypothesis, the corona is heated by the Joule heat of the electric currents set up by variations of the magnetic field in the

corona. This hypothesis is so far not amenable to quantitative verification, since neither the strength of the magnetic field in the corona nor the rate of its time variation in different regions are known exactly. Naturally, in general, variations of the magnetic field caused by "adhering" of the coronal gas to the magnetic lines of force (the corona is an excellent conductor due to the total ionization of its hydrogen and helium) and by the enormous self-induction of the corona, need not induce substantial electric currents. A concentration of the magnetic force lines will compress the corona and heat it briefly, but in places where the magnetic lines of force will compress the corona and heat it briefly, but in places where the magnetic field strength decreases, the gas can also cool. Finally, the coronal magnetic field cannot be held responsible for the heating of the corona as a whole, since this field may fluctuate only due to turbulence. In such a case, the magnetic field must be regarded as turbulent, with randomly twisted lines of force. The mean strength of such a field will be determined by the equality of the magnetic energy and kinetic energy (that is, the energy of turbulent mixing) in the corona:

$$\frac{\varrho v^2}{2} = \frac{H^2}{8\pi}, \qquad (11)$$

where ϱ is the density. The magnetic field therefore cannot heat the corona any more effectively than the decay of turbulent streams does. Actually, the magnetic field merely appears to set up a certain characteristic magnetic viscosity accelerating the decay of the mechanical motions.

The coronal magnetic-field variations are produced by an external source. The enhancement of the coronal emission and the rise in temperature (on the average) above active solar regions may indicate that the corona is heated in some way by variations of the magnetic fields of the sunspots and by the briefly existing fields of the chromospheric flares. It is possible that the heating mechanism can be explained by shock waves which arise when H varies rapidly and the conductor (the coronal gas) is moved. A theory describing these processes has not yet been developed.

The second problem of coronal theory, that of the dissipation [or evaporation] of gas from the corona, is essentially still being formulated, although contrary to the first problem (coronal heating) the physical meaning of the evaporation is entirely clear.

Formerly this problem was in general related to the volatization of light gases from planetary atmospheres and to the possibility of the retention of appreciable atmospheres around low-mass planets. Evaporation from the corona was first discussed by V.A.Krat /9/, who considered it in conjunction with the problem of the net positive electric charge of the corona. The problem can be divided into two parts.

1) Determining the relation between the ejection of variously charged particles by the sun and the intensity of the electrostatic field.

2) Determining the rate of evaporation (the corpuscular flux) of gas from the corona.

Although the coronal magnetic field has a decisive influence on the motion of the particles of the electric plasma, the complexity of the problem and the uncertainty of what was known about the magnetic-field intensity led researchers to neglect this factor and to take the coronal magnetic-field intensity as zero. It is only in this case that a solution can be obtained

by gas-kinetic methods, by generalizing these methods to the motion of plasma particles. Therefore, here we will deal with electrostatic and gravitational fields only. Since the mean free paths of the particles are invariably much less than the dimensions of the coronal inhomogeneities, we need not assume a spherically symmetric corona and calculations can be made for different regions of the solar surface. A particle escapes from the sun only if its velocity exceeds the escape velocity v_e:

$$v_e^2 = 2gR, \tag{12}$$

where g is the acceleration of the external field (the resultant of the gravitational plus the electrostatic field), and R is the distance from the center of the sun. In order to prevent a continual charge accumulation at the boundary of the corona, the total flux of particles evaporating from the corona must be quasineutral. In other words, the outgoing fluxes of differently charged particles should be equal to one another. Let us assume that the energy distribution of the electrons is given by the function $S'(\varepsilon)$ (where ε is the energy) and that of the protons is given by $S(\varepsilon)$. Obviously, if the electron temperature is higher than the proton temperature, then $S'(\varepsilon) > S(\varepsilon)$. The weight of the escaping protons will be denoted as $m_p g_p$ (where g_p is the acceleration of the protons) and that of the electrons as $m_e g_e$ (where g_e is the acceleration of the electrons). If $1-\beta$ is the fraction of the acceleration of gravity (g_0) which is compensated for by the positive charge of the proton, then we have

$$\begin{aligned} m_p g_p &= m_p g_0 (1-\beta), \\ m_e g_e &= m_p g_0 \beta, \end{aligned}$$

or

$$\left.\begin{aligned} g_p &= g_0 (1-\beta), \\ g_e &= \frac{m_p}{m_e} g_0 \beta; \end{aligned}\right\} \tag{13}$$

here $\beta > 0$ indicates that some positive electrostatic acceleration exists. Within a hydrogen plasma (which is a good approximation of the solar atmosphere), $\beta = 0.5$. The equality of the weights of all the particles makes them indistinguishable from one another and ensures quasineutrality of the plasma.

Let us now consider some spherical surface around the sun through which the flux of evaporating particles passes. The radius of this surface can be varied from some relatively small value R_1 for which the highest number of particles is evaporated by every volume element of the gas (the so-called escape layer) up to arbitrarily large values of R. For our purposes it suffices to take R close to the radius R_1 of the escape layer ($R_1 \approx 2R_\odot$). The corresponding spherical surface is called the escape surface or evaporation surface. The particles passing through this surface no longer lose energy due to collisions, and thus they escape from the sun for good. Let us denote the limiting electron and proton energies needed to overcome the potential barrier of a star, and thus corresponding to the escape velocities, as ε_e and ε_p, respectively. The equality of the corpuscular fluxes escaping through 1 cm^2 of the escape surface then may be expressed as

$$\int_{\varepsilon_p}^{\infty} S(\varepsilon)\, v(\varepsilon)\, d\varepsilon = \int_{\varepsilon_e}^{\infty} S'(\varepsilon)\, v'(\varepsilon)\, d\varepsilon, \tag{14}$$

where

$$v = \sqrt{\frac{\varepsilon kT}{m_p}}, \quad v' = \sqrt{\frac{\varepsilon kT}{m_e}}, \tag{15}$$

$$\varepsilon_p = \frac{m_p v_{e,p}}{kT}, \quad \varepsilon_e = \frac{m_e v_{e,e}}{kT} \tag{16}$$

($v_{e,p}$ and $v_{e,e}$ being the escape velocities for protons and electrons, respectively).

When (15) is substituted into (14), we obtain

$$\left(\frac{1}{m_p}\right)^{1/2} \int_{\varepsilon_p}^{\infty} S(\varepsilon)\, \varepsilon^{1/2}\, d\varepsilon = \left(\frac{1}{m_e}\right)^{1/2} \int_{\varepsilon_e}^{\infty} S'(\varepsilon)\, \varepsilon^{1/2}\, d\varepsilon, \tag{17}$$

from which

$$\left(\frac{m_p}{m_e}\right)^{1/2} = \frac{\int_{\varepsilon_p}^{\infty} S(\varepsilon)\, \varepsilon^{1/2}\, d\varepsilon}{\int_{\varepsilon_e}^{\infty} S'(\varepsilon)\, \varepsilon^{1/2} d\varepsilon} \tag{18}$$

If the electron gas and the proton gas have the same temperature (this was shown by Shklovskii to be true for the coronal gas on the average), then $S(\varepsilon) = S'(\varepsilon)$. Thus, condition (18) will be satisfied only if

$$\varepsilon_p < \varepsilon_e \tag{19}$$

or

$$m_p v_{p,e}^2 < m_e v_{r,e}^2 \tag{20}$$

and furthermore if

$$m_p g_p < m_e g_e,$$

$$1 < \frac{\beta}{1-\beta},$$

that is, if

$$\beta > 0.5. \tag{21}$$

If the coronal gas has an equilibrium (Maxwell) distribution, then we have $\beta = 0.58$ for $T \approx 10^6$ deg. For $T_e > T_p$ (the corona is heated by currents, and the electron temperature is somewhat higher than the proton temperature), inequality (19) becomes even sharper, since in this case $S'(\varepsilon) > S(\varepsilon)$ for ε close to ε_p and ε_e. Thus, β may prove to be close to unity. However, as Pikel'ner correctly observed, for $\beta \approx 1$ evaporation becomes a decisive factor in the cooling of the corona, and it may even happen that $T_p > T_e$ with a subsequent reduction of β.

Although in general β is somewhat higher than 0.5, it does not differ greatly from its equilibrium value within a plasma. If we take $\beta \approx 0.6$, we will obtain a positive charge of $3 \cdot 10^{11}$ electrostatic units for the sun. When β drops below the value prescribed by condition (18), the electron flux becomes greater than the proton flux and the sun's charge will increase; when β rises, on the other hand, the enhancement of the proton flux will lower the charge. Magnetic-field effects cannot cause qualitative changes in this picture of the evaporation, since, regardless of the exact form of the magnetic field, for $\beta = 0$ the flux of escaping electrons will be greater than the flux of escaping protons. As a result, some steady state is reached in which $0 < \beta < 1$.

However, the situation is entirely different if we wish to calculate the total flux of escaping particles and then to estimate the loss of solar mass due to the evaporation. Here we cannot manage without calculating the effect of the coronal magnetic field on the evaporation. Even a weak magnetic field with H values of some tenths or some hundredths of a gauss will drive the evaporating corpuscules right along the magnetic lines of force, thus impeding their eventual escape. No particles having ε close to ε_p and ε_v will escape, and the potential barrier will rise sharply. Calculations of the total flux of evaporating particles, carried out by V. A. Krat neglecting the magnetic field and independently by Pikel'ner /9/, gave estimates which agreed in value; these authors calculated that the sun loses about 10^{17} g/year of its mass by evaporation, which is some hundreds of times smaller than the mass losses due to radiation. The true evaporating flux must be even less than this figure.

Geoactive particles can be ejected from the sun only in the form of quasineutral clusters of matter, since even a slight departure from quasineutrality due to an excess of particles of one particular charge in the cluster will result in a high electrostatic acceleration preventing charge separation and tending to stop the motion of the cluster; this will either lead to quasineutrality of the cluster (due to the escape from it of the excess particles disturbing quasineutrality) or else it will make the cluster stop. Even the presence of slight charges is not permissible, since it would greatly increase the influence of the coronal magnetic field on the motion of the clusters.

In conclusion, let us consider one more particular problem of coronal theory, pertaining to the structure of the corona. The cooler "red" regions in the corona adjoin the hot "green" regions, but the gases from the different regions do not mix together, in spite of the substantial pressure gradients between them. Shklovskii has given the correct explanation of this fact. He showed that the presence of the coronal magnetic field greatly lowers the heat conductivity and electrical conductivity of the coronal gas in a direction perpendicular to the magnetic lines of force. This magnetic effect already becomes noticeable for H values greater than 10^{-7} gauss. Since the turbulent magnetic field in the corona is actually much stronger than this, there will be almost no mixing of gases of different temperatures via heat conduction. This appears to explain how coronal rays can exist for some weeks.

Considerable importance was previously attached to the outer coronal rays, which were regarded as corpuscular streams ejected by the sun. If a ray is assumed to be a gas streamer, then we must assume that the condition

$$v(\varrho)\, W^2 n_e = \text{const} \qquad (22)$$

is satisfied, where W is the ray thickness, $v(\varrho)$ is the streamer velocity for a given ϱ, and n_e is the electron concentration (proportional to the streamer density). Since for a streamer ejected from the sun, $v(\varrho)$ can only decrease with height, $W^2 n_e$ must rise as ϱ increases. Moreover, van de Hulst has shown that photometric data on coronal rays give a drop in $W^2 n_e$ with increasing ϱ. Moreover, if the coronal rays represent corpuscular streams leaving the sun, then in the inner corona the gas velocities in these streamers must be greater than the escape velocity for neutral bodies, that is, greater than about 600 km/sec. Actually, in the corona even considerably lower velocities of ordered motion than this are not observed.

REFERENCES FOR CHAPTER VII

1. KRAT, V.A. — Izvestiya Glavnoi astronomicheskoi observatorii, No.142. 1949.
2. VAN DE HULST. H.C. Chapter V of "The Sun," edited by G.Kuiper. — Izdatel'stvo inostrannoi literatury. 1957. [Russian Translation from English.]
3. TIKHOV, G.A. —Izvestiya Glavnoi astronomicheskoi observatorii, No.110. 1931; Trudy ekspeditsii po nablyudeniyu solnechnogo zatmeniya 1936 g., Vol.2. 1949.
4. BUGOSLAVSKAYA, E.Ya. — Trudy Gosudarstvennogo astronomicheskogo instituta im. Shternberga, Vol.12. 1949.
5. CSADA, I.K. — Acta Phys. Chem. Szeged 4(3—4): 77. 1958.
6. SHKLOVSKII, I.S. Solnechnaya korona (The Solar Corona). — Gostekhizdat. 1951.
7. WALDMEIER, M. Die Sonnenkorona, Bd.2. — Basel. 1957.
8. TRELLIS. Theses Univ. — Paris. 1956.
9. PIKEL'NER, S.B. — Izvestiya Krymskoi astrofizicheskoi observatorii, 5: 34. 1950.

M. N. GNEVYSHEV

N 67-15492

Chapter VIII

THE SOLAR SURVEY

§50. A definition of the term "Solar Survey." Various phenomena in the solar atmosphere, such as sunspots, faculae, prominences, and coronal formations, are quite unstable and they change very rapidly. Consequently, in most cases isolated observations at any particular time are of no great significance, and it is mainly interesting to study the development of these phenomena from their birth to their final decay, together with the accompanying variations in the relevant physical characteristics.

Solar phenomena are closely related to one another, so that no one phenomenon should be studied without considering the simultaneous variations taking place in the others. Therefore, observations of solar phenomena should be both continuous and comprehensive. The observations must be made systematically, and the phenomena occurring over the entire surface and in all the observable layers of the sun should be followed simultaneously.

The tendency of scientific progress is such that requirements like these become constantly more demanding. Previously, observations yielding one photograph daily were considered satisfactory, but now continuous motion-picture recordings of the phenomena occurring in various layers of the solar atmosphere are made. This applies even to comparatively slowly changing phenomena, such as sunspots.

Naturally, continuous round-the-clock observations are unfeasible in any one location, and several observatories must cooperate; these observatories should be distributed quite densely according to longitude, so that interruptions due to poor weather or nightfall in some observatories will be compensated for by regular observations in others. In order to secure such a reciprocal compensation, the participating observatories must be equipped with identical instruments and must all follow closely coordinated observation programs using the same observation methods. The totality of measures undertaken in the system of observatories, which together are designed to obtain the most detailed continuous series of observations possible, have become known as the Continuous Solar Survey, or in short, The Solar Survey.

A very important stimulating factor in the development of the Solar Survey has been solar-terrestrial interactions: solar phenomena influence the earth's atmosphere, causing polar auroras, magnetic storms, ionospheric changes leading to radio fadeout, etc. The concomitant applied problems, particularly those connected with ensuring uninterrupted radio communications, also require continuous and comprehensive observations of solar phenomena in order to collect data suitable for comparison with the relevant phenomena in the earth's atmosphere and to predict the latter.

The Solar Survey recently received considerable impetus in connection with the International Geophysical Year and the Year of Geophysical Cooperation (1957 - 1959), during which the number of observatories participating in this project was increased considerably. The equipment was brought up to date, and comprehensive obligatory programs were prepared for all the observatories participating in the Solar Survey.

Initially, the Solar Survey was made using small, comparatively primitive instruments, since there were very few large instruments which had adequate resolving power and which produced sufficiently large images. This had, of course, an unfavorable influence on the quality of the data obtained by the Solar Survey, particularly where chromospheric phenomena were concerned. Therefore, historically we can divide solar observations into those of very high quality but which were made unsystematically, at random moments, and those of comparatively low accuracy but which were made systematically and continuously. It is the latter group that constituted the original Solar Survey.

Recently, however, two things became clear, On the one hand, the regularities which could be observed with small instruments had already been discovered, while, on the other hand, isolated, high-quality photographs taken using large instruments were of little use, since they could not provide a picture of the development of the phenomena. The principal organizational problem of the modern Solar Survey is therefore to ensure continuous observations (often even to use motion pictures) over a quite dense network of observatories equipped with high-resolution instruments.

In order for this collective project to be a success, the observatories taking part in the Solar Survey must establish a constant and rapid exchange of information. This condition imposes rigid requirements on the processing priority of the observational data (for ordinary astronomical observations, the processing stage lags behind the observation considerably, sometimes even a few years). The great urgency of the data processing and the information gathering is also due to the fact that, as mentioned above, data pertaining to changes in the solar atmosphere are essential for the solution of various applied problems relating to phenomena in the terrestrial atmosphere.

Observatories are joined for the Solar Survey not only on a national scale, but also internationally, under the direction of the International Astronomical Union. The following observatories are the most prominent contributors to the Solar Survey. They are distinguished by the homogeneity and length of their solar-observation series: Zurich, Greenwich, Meudon, The Fraunhofer Institute (Freiburg, West Germany), and Mount Wilson; international surveys providing extra-eclipse observations of the corona and the chromospheric flares also make invaluable contributions to the Solar Survey. Finally, in the USSR, the Solar Survey is organized and monitored by the Solar-Research Commission for the Astronomical Council of the USSR Academy of Sciences. A detailed description of the Solar Survey in the Soviet Union is given in /1/.

Since systematic and quite frequent measurements of all the possible characteristics of all the solar phenomena cannot be made in practice, therefore systematic sets of data are compiled only for the measurements of certain individual characteristics, known as solar-activity indexes. These indexes serve as indicators of the general state of the solar

activity. Solar-activity indexes enable us to become oriented with respect to the vast amount of data at hand, and they indicate those interesting cases in which the photographs must be measured with a maximum possible carefulness and completeness. Furthermore, solar-activity indexes are used for preliminary comparisons both between different solar phenomena and between solar phenomena and geophysical processes.

The most common solar-activity indexes are the spot numbers (Wolf numbers), the sunspot areas, the sunspot magnetic-field intensity, the facular area, the area and intensity of hydrogen and ionized-calcium flocculi, the number, area, shape, and intensity of hydrogen prominences, the intensity of coronal emission lines, and the radio-emission intensity at various wavelengths.

Of great significance is the survey of chromospheric flares; these phenomena are very short-lived, so that round-the-clock observations are especially necessary.

In addition to these indexes, which are observed and published regularly, other indexes introduced by various observers are published from time to time. However, none of these have become subjects of constant observation, mainly due to the complexity of their determination and also for other reasons. Of these special indexes, let us cite the ratio of intensities for the H_ε line of hydrogen and the H line of ionized calcium in prominences; this ratio was determined by Perepelkin for the period from 1928 to 1935, and it served as an indicator of the intensity of the far ultraviolet radiation of the sun. Other important indexes are the intensity of coronal glow in the continuum and the intensity of the general magnetic field of the sun (the latter is not yet being measured systematically).

§51. Photospheric solar-activity indexes. The first observations that can be regarded as pertaining to the Solar Survey were made soon after the invention of the telescope, and they amounted to just sunspot counting. These observations were fairly sporadic at first, and each series terminated at the end of the active period of the observer's life. Then, as the number of observers and observatories increased, the accumulation of more complete data became possible.

Systematic determinations of the spot numbers were begun in 1847 by R. Wolf at the Zurich Observatory (see §18), and Wolf continued his observations for 45 years. Since that time these observations have been carried on continuously according to a single program. Subsequently, the Zurich series were expanded to include the data of other observatories and of individual observers. Moreover, observations from previous years were collected, analyzed, and reduced to the Zurich system, so that the series of sunspot numbers was extended back to the year 1700 (yearly averages). Also, the years of sunspot maxima and minima have been recorded back to 1610; these data, of course, are not as reliable as the more recent measurements.

The Wolf numbers have also been derived according to Soviet observations /2/, and a comparison of these numbers with the Zurich series shows excellent agreement between the two sets of data /3/. In general, sunspot numbers represent a very crude and subjective index of solar activity, since these numbers depend greatly on observation conditions and also on the particular grouping of sunspots preferred by the individual observer; the latter factor is especially influential when spots are abundant and when spot

groups are distributed densely over the surface of the sun, so that differences of opinion may arise as to the exact grouping of the different spots. On the other hand, the great advantage of the Wolf numbers is the simplicity of their determination and the fact that (due to their simplicity) they can be computed over several centuries. For these reasons, it will be useful to continue computing the Wolf numbers in the future, even though at present there are more precise indexes characterizing the appearance of sunspots.

In order to correlate sunspots with geophysical phenomena, it is advisable to calculate the Wolf numbers not for the entire solar disk, but rather just for the central zone within a circle with a radius of one half the solar radius. These numbers have also been computed for many years into the past.

A more precise and objective index characterizing the sunspot activity is the spot area, which is generally expressed in millionths of the solar hemisphere. The corresponding observations are much more complex than obtaining the sunspot numbers is, since to determine the area of a spot it is necessary to measure its coordinates, then to measure the spot area on a photograph or screen, and finally, making allowance for the distance of the spot from the apparent center of the solar disk, to correct the measured area for the spherical projection of the sun. These measurements and reductions become progressively less accurate for sunspots located more toward the limb of the solar disk. With the approach to the limb, the accuracy of spot-area measurements drops, because the projected image of a spot becomes smaller and the image vibrations increase; furthermore, it should be remembered that spots are not actually plane formations. For these reasons, small spots disappear before going beyond the solar limb, and the projection correction based on multiplying by the secant of the angular distance of the spot from the center of the apparent disk becomes inadequate. This question has been investigated by various authors /4/. As a result, a special visibility function has been introduced, with which it is possible to correct the observational data so that, when averaged over zones at different distances from the center of the apparent disk, they do not depend on this distance. The influence which the observation conditions have on the visibility of spots near the solar limb leads to minor seasonal variations in the averaged data on spot numbers and areas. Naturally, a statistical processing of the data will eliminate these variations, since they are unrelated to the solar activity.

In 1907 Maunder /5/ established that there is a slight difference between the number and size of sunspots at the eastern and western edges of the solar disk. This problem was subsequently discussed in numerous works, and very interesting investigations of the subject are now being made by Dezse /6/ at the Budapest Observatory. The most probable explanation of this phenomenon is that a spot axis is alternately tilted in the direction of solar rotation and in the opposite direction, depending on the age of the spot group. Thus, since spots constitute funnellike depressions, the conditions of their visibility will differ at the eastern and western limbs.

In order to characterize sunspot activity, the total area of all the sunspots on a given day is calculated. Here, depending on the purpose of the observations, the count is made either for the entire disk or just for the central zone. Of course, if small spots are missed due to atmospheric factors or to instrumental imperfections, this does not influence the total spot area much,

whereas for the calculation of spot numbers small sunspots have the same weight as large ones.

Sometimes, when the spot areas are to be correlated with terrestrial phenomena, the index is not corrected for the curvature of the solar surface. Then, the spot areas are expressed in millionths of the solar disk. Naturally, the application of such an index implies that the radiative emission influencing the earth's atmosphere is proportional to the spot area and obeys Lambert's law. This is a permissible assumption for the luminous radiation but is definitely false for the corpuscular radiation emitted in narrow beams (see below) and producing magnetic disturbances, polar auroras, and special ionospheric disturbances in the earth's atmosphere.

Regular measurements of the spot areas and coordinates were made by Spörer /7/ between the years 1861 and 1868. Since 1874 such observations have been regularly made at the Greenwich Observatory, and for this purpose special photoheliographs were installed at the Greenwich, Kodaikanal (South India), and Cape (South Africa) stations. The photographs obtained at these observatories and processed at one place provide a continuous series of spot areas and coordinates which is distinguished by high accuracy and homogeneity.

Starting in 1932, similar data were obtained in the USSR on the basis of observations made at Simeis, Tashkent, Kiev, Irkutsk, Odessa, and Kazan. Subsequently, the Soviet photospheric survey was restricted to observatories equipped with the standard meniscus photoheliograph of Maksutov's system; these observatories are: the Mountain Astronomical Station of the Pulkovo Observatory near Kislovodsk (functioning since 1948), the Tashkent, Kiev, and Irkutsk Observatories, and the Far-East Station at Ussuriisk.

The data of the Soviet Solar Survey are collected into solar-activity catalogs published by Pulkovo Observatory; since the end of 1954, the preliminary data have been listed in the bulletin "Solnechnye Dannye," published jointly by the Solar-Research Commission and Pulkovo Observatory. In these publications, the data of all the observatories are reduced to the photographic-observation system of the Mountain Astronomical Station. A comparison of these data with the Greenwich series /8/ shows that the numerical values of all the characteristics are virtually identical.

Another solar-activity index is the facular area. Before the introduction of the photographic method of observation, the area occupied by faculae was estimated crudely in terms of the angle between lines drawn from the center of the apparent solar disk to the end points of the facula. Photographic techniques, however, have made it possible to determine the areas and positions of faculae in the same way as the areas and positions of sunspots, and to express these quantities in the same units. However, the quantitative index or the facular area should be applied only after averaging over a period of not less than one solar rotation, since faculae become visible only near the limb of the apparent solar disk; thus, the total area of faculae visible on any given day is not representative of the actual facular area.

Subsequently it was established that calcium flocculi in the K_3 and H_3 lines coincide exactly with the photospheric faculae in size, position, and shape; the flocculi have the advantage of being visible from any part of the disk. Therefore, indexes based on observations of calcium flocculi are equivalent to facular indexes, and they are more comprehensive in that they characterize each single facula.

So-called synoptic charts also constitute a very convenient and illustrative means of representing solar data (Figure 58). These charts show the surface of the sun in cylindrical or Mercator projection. Vertical lines indicate the solar meridians, reckoned from Carrington's meridian at intervals of 10° longitude (in Figure 58 the longitudes are indicated along the top of the chart). Horizontal lines show the parallels, separated by intervals of 10° latitude (the latitudes are indicated along the sides of the chart).

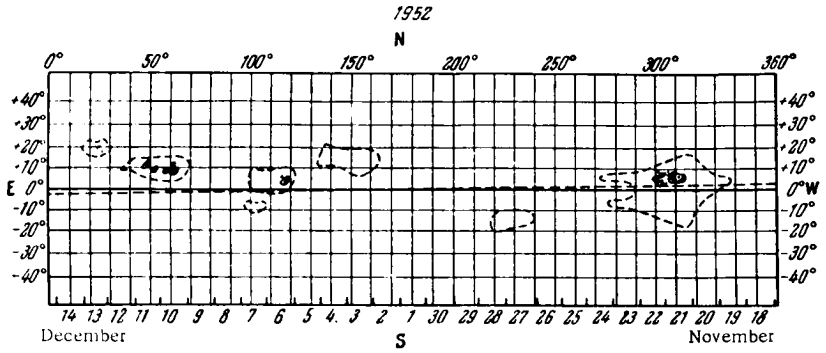


FIGURE 58

The sun completes a synodic rotation about its axis in about 27 days. During this period every solar meridian faces the earth at some time, and systematic observations can record all the solar details, so that their shapes and positions can be recorded on the chart. At the bottom of the chart are shown the times during which a given meridian faces the earth. Each synoptic chart thus represents the solar surface, spread out in projection, for a given solar rotation. The disadvantage of these charts is that they do not show the changes in solar formations occurring during the given solar rotation. Nevertheless, they do give an adequate picture of the distribution of the principal centers of solar activity. In addition, a comparison of charts pertaining to successive rotations indicates the principal changes in activity.

The dashed "straight" line shows the position of the earth's projection onto the face of the sun. The intersection of this line with each meridian in turn gives the point on the sun which was at the center of the apparent disk on the day marked at the bottom of the chart. Synoptic charts make it easy to solve any geometrical problems which come up when solar and terrestrial phenomena are compared.

Faculae, which are often quite extensive in size, are plotted onto synoptic charts part by part, as they emerge from beyond the eastern limb or as they disappear at the western limb. A synoptic chart thus eventually gives the complete true form of a facula, and thus the disadvantage that faculae are visible at the edge of the sun only is compensated for. A regular construction and publication of photospheric synoptic charts has been carried out by the Zurich Observatory /9/ since 1926. Also, since 1932, similar photospheric synoptic charts compiled from Soviet observations have been published in the Pulkovo catalog of solar activity.

Since synoptic charts do not show the day-to-day changes in solar formations, special charts are now being published giving the daily appearance of the solar surface; the sunspots, calcium flocculi, filaments, and prominences appearing on photoheliograms and spectroheliograms are indicated with great precision on these charts, and the true shapes and sizes are given. Furthermore, these charts give the intensity distributions of the green and red coronal lines around the sun, according to extra-eclipse coronagraph observations. Figure 59 shows one such daily chart from the bulletin "Solnechnye Dannye," in which they have been published since 1954. Analogous charts are also published by the Meudon Observatory and by the Fraunhofer Institute.

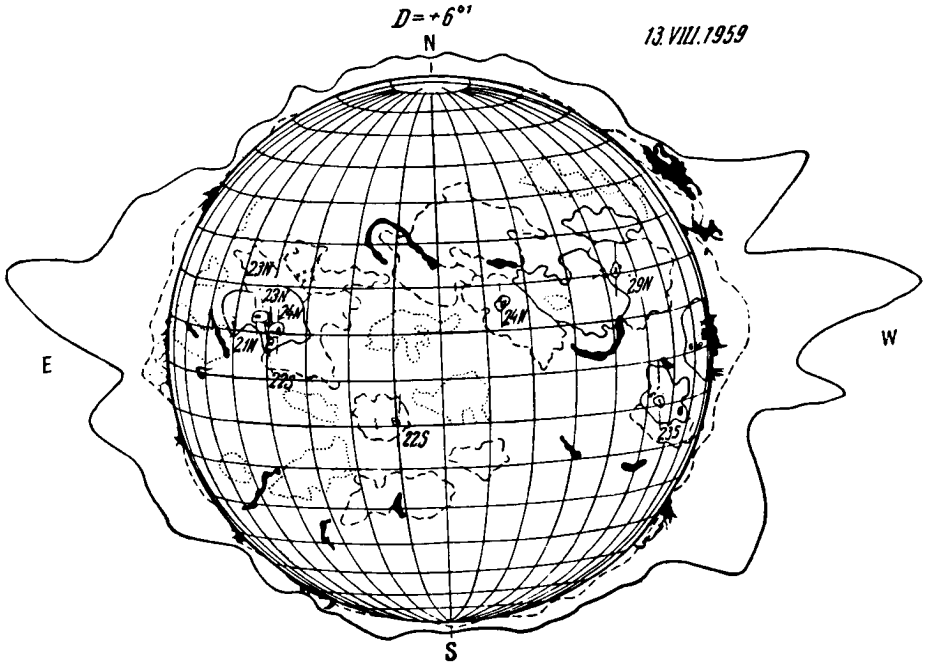


FIGURE 59

§52. Chromospheric solar-activity indexes. Some common chromospheric indexes are: the coordinates, areas, and intensities of flocculi (mainly of ionized calcium), and the areas, shapes, positions, and intensities of hydrogen prominences, filaments, and ejections. The intensities of sunspot magnetic fields also constitute a chromospheric index, since they are determined from the Zeeman splitting of chromospheric lines.

Calcium and hydrogen flocculi were characterized in former times by very crude, visually estimated indexes; the area of all the flocculi was considered together, along with their general compactness and intensity. These estimates were obtained by eye, by comparing the picture obtained through a spectrohelioscope with appropriate standards, namely with spectroheliograms taken on days of different activities. Such estimates

were published in the quarterly solar-activity bulletin of the International Astronomical Union from 1917 to 1940 /10/.

In 1921, d'Azambuja /11/ started regular observations of chromospheric formations with the large spectroheliograph at Meudon. Subsequently, the Mount Wilson and Kodaikanal Observatories began to take part in this project with similar instruments. The synoptic charts of the chromosphere based on these observations and published yearly by the Meudon Observatory have so far found no equals as far as completeness and accuracy are concerned. (The same charts are published monthly in "L'Astronomie.") Figure 60 shows one of these charts. The chromospheric charts are prepared by the same method as the photospheric charts described in the preceding section. They give drawings of the calcium flocculi in which the shapes and dimensions are shown.

The density of cross-hatching indicates the intensity of the flocculi according to a five-point scale. The charts also show filaments according to their true shapes; sunspots are shown schematically as circles whose areas are proportional to the spot area. The varying darkness of the different parts of filaments indicates the filament lifetime, the older parts being darker. Chromospheric charts are accompanied by tables listing the numerical characteristics of the filaments shown in the charts.

Another important publication is the Coimbra Observatory Albums, which list daily spectroheliograms obtained with a large spectroheliograph in the K_1 and K_3 lines of ionized calcium /12/.

Lately, spectroheliograms of very high quality have been obtained with the large solar telescope of the Crimean Astrophysical Observatory, and also at the Mountain Astronomical Station. Small-size photographs are also obtained at the Khar'kov, Abastumani, and Simeis Observatories.

Data on prominences are published regularly by the Zurich Observatory /9/ in the form of tables. The prominence area is given for each day of observation, for each 5° of position angle. The Pulkovo solar-activity catalog for the years 1933 through 1956 listed the total prominence areas over intervals 30 latitude degrees in width, according to observations at Simeis, Tashkent, Abastumani, and the Mountain Astronomical Station.

Since 1957 more complete data have been published on the basis of observations with extra-eclipse coronagraphs at the Mountain Astronomical Station of Pulkovo Observatory, the Mountain Station of the Astrophysical Institute at Alma Ata, and the Crimean Astrophysical Observatory. These observations are used to record the prominences, by indicating their positions, shapes, and scales, in the daily charts of the bulletin "Solnechnye Dannye"; moreover, the Pulkovo catalog of solar activity lists in tables numerical data on the positions, areas, heights, and intensities of all prominences.

Finally, fragmentary prominence series of varying length have been obtained by individual observers. Of these we should mention the observations of line-of-sight prominence velocities carried out at Pulkovo /13/.

Considerable progress in the investigation of prominences was achieved when slow-motion moving pictures began to be used. At present this technique is widely applied, so that it is possible to refer to a motion-picture prominence survey. The motion-picture photography of prominences is much used at Mount Wilson (since 1936), at the McMath-Hulbert Observatory (since 1937), at the High-Altitude Observatory in Climax (since 1942), and

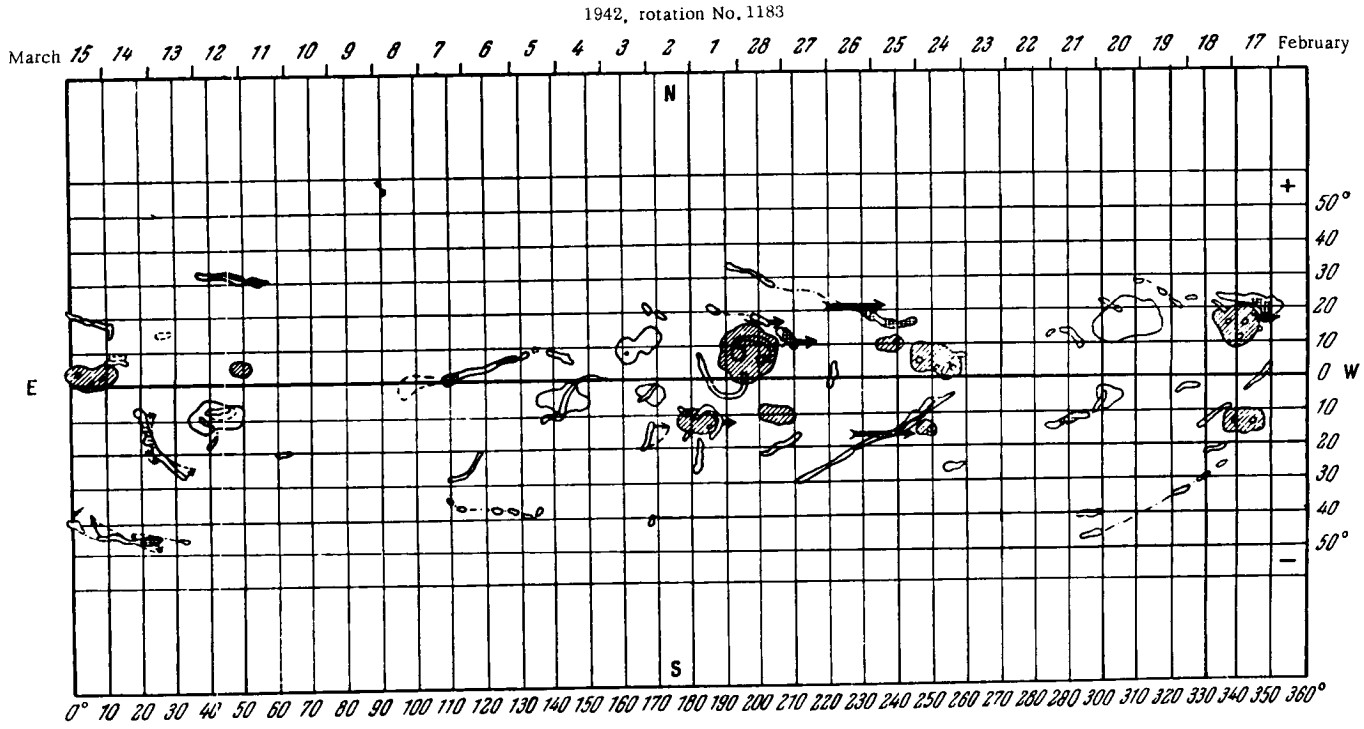

1942, rotation No. 1183
March 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 28 27 26 25 24 23 22 21 20 19 18 17 February
N
S
E
W
+
−
50°
40
30
20
10
0
10
20
30
40
50°
0° 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360°

FIGURE 60

at the Sacramento Peak Station (since 1950). In the USSR, the motion-picture photography of prominences has been employed widely at the Crimean Astrophysical Observatory since 1950.

Of considerable interest also are observations of chromospheric flares, since they throw light on the nature of processes going on in active regions of the sun and in addition reveal a definite correlation with phenomena in the upper layers of the earth's atmosphere.

Until recently, visual observations were the only source of data on chromospheric flares; many observatories, each equipped with spectrohelioscopes or monochromatic filters, were engaged in these observations. The results were (and are) then published systematically in the IAU quarterly bulletin on solar activity. In addition, flare observations by Soviet observatories are published in the Pulkovo catalog and in the bulletin "Solnechnye Dannye." However, the application of motion-picture photography is now also very useful for observations of chromospheric flares /14/. In the USSR flares have been filmed through chromospheric telescopes since the beginning of the International Geophysical Year. Chromospheric flares are very closely related to certain electrical phenomena in the earth's atmosphere, and the latter can be used as indicators signifying the appearance of a flare. The most convenient of these indicators is the rise in the intensities and frequencies of atmospheric discharges at 27 kc during a chromospheric flare. Since these atmospheric discharges are recorded by continuous automatic devices, such observations will detect all the flares, regardless of weather conditions; such recordings are much easier than optical ones, which require the presence of an observer. The appearance of large bursts over a wide frequency range of the solar radio emission (20 to 33,000 Mc) constitutes another excellent indicator of chromospheric flares.

The magnetic fields of sunspots have been observed systematically at Mount Wilson Observatory since 1917. The data for 1917 through 1924 were published in /15/, and subsequent data have been published regularly in /16/. These publications quote for each spot group the magnetic polarity and the magnetic-field strength (in gausses). In the USSR, analogous observations are being carried out at the Institute of Terrestrial Magnetism, at the Crimean Astrophysical Observatory, at Pulkovo Observatory, and (since the beginning of the International Geophysical Year) at the Mountain Astronomical Station of Pulkovo Observatory, near Kislovodsk.

§53. Coronal solar-activity indexes. Systematic observations of the corona became possible only with the introduction of extra-eclipse observations; the latter were first accomplished by Lyot at Pic du Midi. The first continuous sets of extra-eclipse coronal observations were accumulated by Waldmeier at Arosa Observatory in Switzerland and published in /17/. These publications show graphically the daily distributions of the green and red coronal lines around the sun at a distance of 1 minute of arc. Later, extra-eclipse observations of the corona were begun at Climax, Kanzelhöhe, Wendelstein, Mount Norikura, and Sacramento Peak, and the data of these observatories (together with those from Pic du Midi and Arosa) are now published in the IAU quarterly bulletin on solar activity. Intensity estimates are made at these observatories in different ways; in most cases 50-point and 40-point relative scales are used. Only at Pic du Midi and Mount Norikura is the intensity expressed in absolute units; here, the total emission

of the given coronal line, relative to the continuum energy at the center of the sun in a 1 Å interval near the coronal line, is given. During the IGY, all observatories switched over to this scale. The measurements are made over intervals of 5° of position angle, reckoned from the north pole of the sun; then they are referred to points in the corona located a distance of 40 arc seconds from the solar limb. The data of Kanzelhöhe and Wendelstein are also published in the Fraunhofer Institute bulletin, "Sonnen Zircular," in tables and in synoptic charts (Figure 61) analogous to the synoptic charts of the photosphere and chromosphere. These charts give the coronal-intensity isolines.

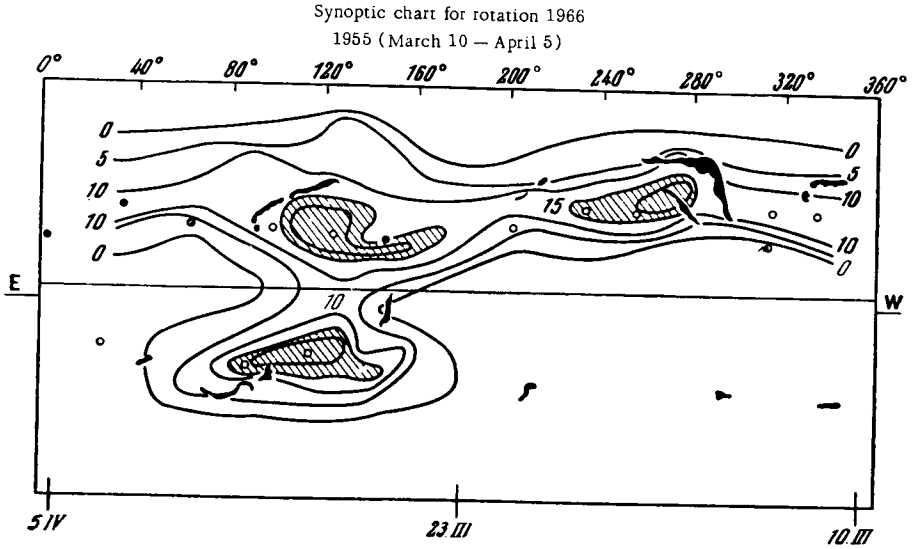


FIGURE 61

In the USSR, systematic extra-eclipse observations of the 5303 Å coronal line were started at the end of 1953 at the Mountain Astronomical Station of Pulkovo Observatory near Kislovodsk. At this same station, from the beginning of 1954, observations were made of the λ 6374 line, and from 1956 of the λ 5694 line and the infrared coronal lines at λλ 10747 and 10798 Å. Furthermore, systematic observations of the λ 5303 and λ 6374 coronal lines have been carried out since 1955 at the High-Altitude Station of the Astrophysical Observatory in Alma Ata. At both these stations the line intensities were also estimated in relative units, but during the IGY they switched over to absolute measurements. The data of the Soviet stations are published in the Pulkovo catalog of solar activity and in the bulletin "Solnechnye Dannye," as well as in the IAU quarterly bulletin on solar activity. Methods of carrying out extra-eclipse observations of coronal lines are described in /18/. Naturally, in order to be useful, the data on coronal lines obtained at different observatories must be reduced to a single system.

§54. Solar radio-emission indexes. Measurements of the intensities and polarizations of the different radio waves emitted by active solar formations now provide a new tool for the observation of active formations in the

chromosphere and corona. The upper chromospheric layers, which form a comparatively dense layer of ionized gas, transmit only radio waves of wavelengths less than a few centimeters (for example, 3-cm waves reach us from the upper chromospheric layers, and 1-cm waves arrive from layers of medium height). Observations at these wavelengths therefore characterize the state of the chromosphere at different depths. Radio waves in the decimenter range and above are not received from regions lying below the corona, so that observations at these wavelengths characterize the state of active regions in the corona. Finally, in cases when the radio waves are divided into ordinary and extraordinary components, polarization measurements give an indication of the magnetic fields.

In view of the preceding, a need was felt for systematic observations of the solar radio emission at various frequencies, since the corresponding data constitute an important addition to the results obtained by optical methods. Such observations are carried out by various institutes and laboratories. The number of stations engaged in this survey is constantly increasing, and the observational techniques and equipment are being improved at a remarkable rate.

Systematic radio data are regularly published in the IAU quarterly bulletin. This publication gives the intensity of solar radio emission at various frequencies from 38 to 9400 Mc for each day of observation, and also the polarization index, calculated from the intensity ratio of two mutually perpendicular vectors (I_R and I_L) as

$$100\,\frac{I_R - I_L}{I_R + I_L}\,.$$

The intensities of these vectors are given in a separate table for each observation day. A polarization index of zero according to this formula obviously corresponds to the absence of polarization.

Also, for each observation day a radio-emission variability index is calculated for various frequencies according to a four-point scale (in this scale, 0 stands for quiet and 3 for highly variable). In addition, a table of outstanding phenomena is compiled. This table lists the starting time of the disturbance, together with its duration, type, intensity, polarization, and a possible identification with optical phenomena.

In the USSR, radio observations are made at Pulkovo Observatory, at the Scientific-Research Radiophysical Institute at Gor'kii, at the Mountain Astronomical Station near Kislovodsk, at Abastumani, at the Crimean and Byurakan Observatories, and at the Irkutsk and Ussuriisk Stations. The Soviet data are published in the bulletin "Solnechnye Dannye."

REFERENCES FOR CHAPTER VIII

1. GNEVYSHEVA, R.S. Nablyudeniya Solntsa (Solar Observations), p.3. — Izdatel'stvo Komiteta Mezhdunarodnogo geofizicheskogo goda. 1959.
2. Solar Activity Catalogs, and also Solnechnye Dannye (Solar Data), a bulletin published by the Main Astronomical Observatory of the USSR Academy of Sciences.
3. GNEVYSHEVA, R.S. — Solnechnye Dannye, 8-9: 22. 1953.
4. ROGGENHAUSEN, M. — Zs. f. Ap. 30: 249. 1952.
5. MAUNDER, A.S.D. — M.N., 67: 451. 1907.

6. DEZSE, L. — Izvestiya Krymskoi astrofizicheskoi observatorii, 16: 208. 1956.
7. SPÖRER, G. — Pub. Ast. Gesellsch. 13: 1. 1874.
8. GNEVYSHEVA, R. S. — Solnechnye Dannye, 8-9: 27. 1953.
9. Astronomische Mitteilungen, Zürich.
10. Quarterly Bulletin on Solar Activity, Zürich.
11. Cartes synoptiques de la chromosphere solaire, Meudon.
12. Anais do Observ. Astr. da Univ. de Coimbra.
13. VYAZANITSYN, V. P. — Izvestiya Glavnoi astronomicheskoi observatorii, 145: 68. 1950.
14. SEVERNYI, A. B. — Izvestiya Krymskoi astrofizicheskoi observatorii, 9: 3. 1952.
15. Magnetic Observations of Sunspots, Pub. Carnegie Inst., 1938.
16. Publ. of the Astron. Soc. of the Pacific.
17. WALDMEIER, M. Die Sonnenkorona. — Bd. I. 1951; Bd. II, 1957. Basel.
18. GNEVYSHEV, M. N. Tekhnika i metodika koronal'nykh nablyudenii (Coronal Observations, Techniques and Procedures), p. 36. — Izdatel'stvo Komiteta Mezhdunarodnogo geofizicheskogo goda. 1959.

N67-15493

A. P. MOLCHANOV

Chapter IX

SOLAR RADIO EMISSION

Radio-astronomical methods of solar research are of very recent origin; the first chance observations were made in 1942, on wavelengths of 4 to 6 meters, by radar stations. The existence of solar radio waves is quite natural, provided we remember that every hot body should emit electromagnetic waves of all frequencies, including waves in the radio range (provided the body can absorb waves in this range). Thus, the first suggestion that the sun emits radio waves was made by Lodge as early as 1894. At about the same time (Scheiner and Wilsing, 1897), attempts were made to detect these waves, but they failed due to the inadequate sensitivity of the equipment. The first radio observations of the sun for scientific purposes were made by Southworth in 1942 - 1943 on wavelengths of 1.25, 3.2, and 10 cm. Since that time, a large amount of observational material has been amassed, and various attempts toward interpretation have been made. It has been found that, in addition to the thermal component, there is also emission which is related to nonequilibrium processes in the solar atmosphere /1-17/.

§55. Methods of observation. Radio-astronomical installations have certain properties which distinguish them from optical equipment and which influence the observation results /13/. These properties must be taken into consideration when comparing the observational data obtained in the different frequency ranges. First, the angular resolving powers of all radio-astronomical installations in current use (including interferometric equipment) are much lower than those of optical instruments. This is quite natural, of course, since the resolving power is determined by the ratio λ/D (where D is the antenna dimension and λ is the wavelength), and since electromagnetic wavelengths in the radio range are several factors of ten higher than optical wavelengths. On the other hand, the dimensions of the antenna arrays (mirrors or multidipole systems) are limited for technical reasons. Thus, it must be kept in mind that a characteristic property of radio equipment, as opposed to optical instruments, is that it allows the operator to tune in easily on relatively narrow transmission bands ($\overline{\Delta f}/f \approx 10^{-3}$), which are then very difficult to broaden. Finally, the radio-astronomical observations which are now being made in the range from millimeter wavelengths up to 15 or 20 meters are, over a very wide range ($5\text{ cm} < \lambda < 3\text{ m}$), practically independent of the state of the earth's atmosphere. *

The first of the properties listed above explains why most of the radio-astronomical observations made until recently have used radiotelescopes with low angular resolving powers. Such instruments cannot be used to

* At shorter wavelengths telluric O_2 and H_2O absorb considerably, while at longer wavelengths the influence of the ionosphere is felt.

study individual features of the solar surface, which is done during optical observations.

This disadvantage, however, has been partially compensated for by radio observations of solar eclipses, when the geometrical relations are such that diffraction phenomena at the lunar limb do not prevent the resolution of different sources located several arc seconds from one another.

TABLE 1

λ, cm	Size and shape	Resolving power for λ_{min} (minutes of arc)	Location
0.8—20	Paraboloid, 22 m diameter	2	Serpukhov (Physics Institute of USSR Academy of Sciences)
3.2—35	Truncated paraboloid, 130 × 3 m	1	Pulkovo (Main Astronomical Observatory)
10	Slit-antenna interferometer in a rectangular wave guide	1	Near Ottawa (Canada)
9.1	Interferometer using 16 × 16 antennas. Base 114 × 114 m	3.1	Stanford (California, USA)
20—20,000	Paraboloid, 66.5 m diameter	15	Jodrell Bank (England)
21	Interferometer using 32 × 32 antennas. Base 378 × 378 m	3	Sydney (Australia)

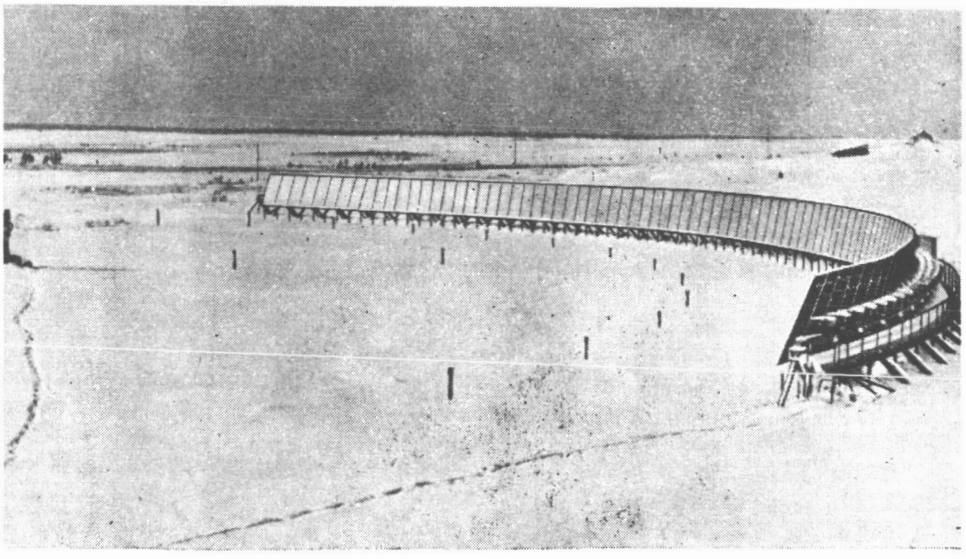


FIGURE 62

Large antennas have been introduced only comparatively recently. The characteristics of the largest existing radiotelescopes (in terms of definite wavelength ranges*) are listed in Table 1. Figures 62 and 63 show photographs of the Pulkovo radiotelescope and of the radiotelescope of the Physics Institute of the USSR Academy of Sciences erected near Serpukhov. The table shows that in resolving power even the largest radio telescopes are much inferior to optical equipment.

FIGURE 63

Direct radio observations of the sun are also supplemented by observations of occultations of radio nebulae by the solar atmosphere. Information on the solar atmosphere is obtained from these observations by measuring the change in the intensity and direction of the radio waves emitted by a radio nebula during its occulation. **

* The errors in the finishing of a reflecting antenna surface cannot exceed $\sim\frac{\lambda}{10}$. One of the principal characteristics of an antenna is thus the shortest wavelength at which it operates.

** Such observations were first staged by Vitkevich, and they led to his discovery of the solar supercorona.

The above-mentioned properties of radio-astronomical observations must always be taken into account when the observational results are being analyzed.

§56. A general description of solar radio emission. According to the modern view, the main part of the solar radio emission with $\lambda < 10$ cm originates in the chromosphere. The longer wavelengths, however, are emitted mostly by the corona. This is explained by the increased absorption of longer wavelengths in the solar atmosphere, together with the general physical law that a heated absorbing medium emits the same frequencies as it absorbs (Kirchhoff's law). According to Milne's criterion (see p. 187 below), the main part of the radiation reaching the observer originates in those layers for which the line-of-sight optical depth is close to 1. Thus, the longer the wavelength, the higher up in the solar atmosphere the emission will have originated.

The radio emission can definitely be separated into undisturbed and disturbed components; the disturbed components can in turn be subdivided into various kinds, the properties of which will be described below. Some general characteristics of the components are listed in Table 2.

TABLE 2

Component	Duration	Wavelength range	Location on the sun
Undisturbed	Continual	All wavelengths	The entire disk
Disturbed, slowly varying	Hours and days	All wavelengths	Regions above sunspots and flocculi
Disturbed, rapidly varying	From tenths of a second to a few hours	All wavelengths	Most bursts appear in regions of chromospheric flares

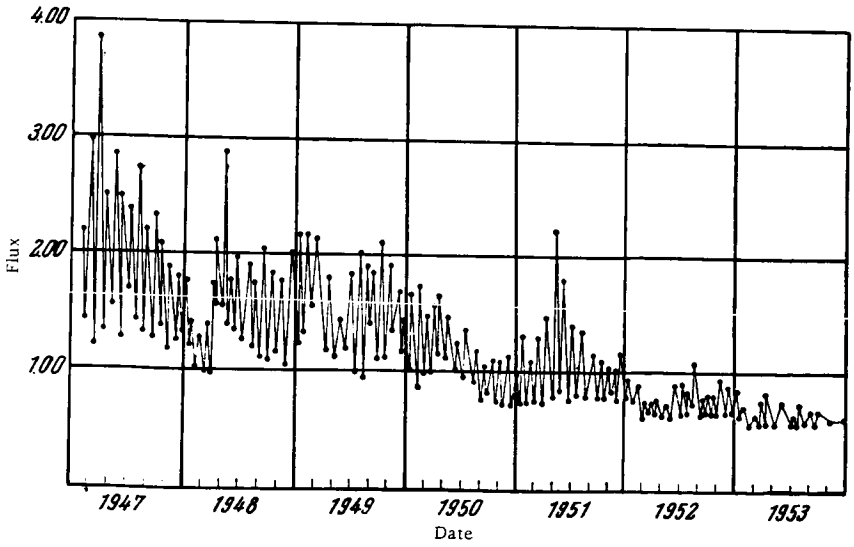


FIGURE 64

The ratio between the first two components is obvious from the graph in Figure 64. This graph shows the maximum and minimum monthly values of the slowly varying component (λ = 10.7 cm) over a period of seven years. The undisturbed component corresponds approximately (see §57) to an envelope drawn through the points of minimum.

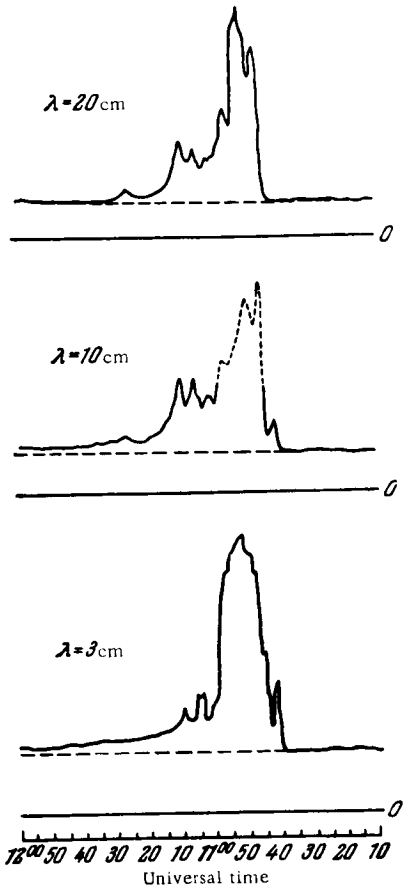


FIGURE 65

The bursts constituting the rapidly varying component differ sharply from the first two components: they are generally related to chromospheric flares, are comparatively short-lived, and they can reach very high intensities. In the short-wave radio range they make the average intensity of solar radio emission several times greater, while in the long-wave range they sometimes increase it several million times. Some examples of bursts are shown in Figure 65.

§57. **The undisturbed component of the solar radio emission.** The distinction between the disturbed and undisturbed components of the solar radio emission is not a completely rigorous one. First, it is very difficult to specify reliably any period during which the sun is actually "quiet." The

solar-activity phenomena recorded by our instruments, such as sunspots, flares, faculae, and prominences, may just be external manifestations of some more general (as yet unstudied) processes, so that they do not completely define an "activity center." However, a complete absence of even just these phenomena is so rarely the case that, in spite of such a limited definition of a "quiet" sun, there is very low probability that it can be observed.

The undisturbed radio component is generally regarded as the thermal emission of the "quiet" sun. Since most of the solar-activity phenomena observed so far have been accompanied by an enhancement of the radio flux, therefore the undisturbed component should apparently correspond to the minimum recorded flux. Thus, it is sometimes determined approximately as the envelope of the minimum values of the measured radio flux, over some fairly long period during years of low solar activity.

Some calculational relations. The thermal radiation component can be estimated theoretically using the transfer equation and the formulas for the absorption of radio waves in an ionized gas (see texts describing the propagation of radio waves in a plasma, for instance /18/.)

The radiation transferred parallel to the observer's line of sight through an area of 1 cm^2 is given by the following equation (Figure 66):

$$\frac{d}{dl} I_\lambda = \chi I_\lambda^* - \chi I_\lambda, \tag{1}$$

where I_λ is the specific intensity of radiation at boundary a of the volume element, and I_λ^* is the specific intensity of radiation at boundary b. For radio frequencies, the latter quantity is defined quite accurately by the Rayleigh-Jeans formula:

$$I_\lambda^* = \frac{2kT_e}{\lambda^2} \frac{\text{erg}}{\text{cm}^2 \text{ steradian}}. \tag{2}$$

In equations (1) and (2), χ is the coefficient of absorption of radio waves, T_e is the electron temperature of the gas, dl is an element of length in the direction of propagation, k is the Boltzmann constant ($k = 1.32 \cdot 10^{-16}$ erg/deg), λ is the wavelength in centimeters, χI_λ^* is the emissivity according to the Kirchhoff-Clausius law for a semitransparent medium, and χI_λ is the attenuation of the specific intensity of radiation passing through the given volume element. The attenuated intensity is thus defined by the formula

$$\left.\begin{aligned} I_\lambda &= I_{0\lambda} e^{-\int_0^{l_0} \chi\, dl} = I_{0\lambda} e^{-\tau}, \\ \tau &= \int_0^{l_0} \chi\, dl. \end{aligned}\right\} \tag{3}$$

The quantity τ is called the optical depth. Since for $\tau = 1$, we have $\frac{I_\lambda}{I_{0\lambda}} \approx 0.37$ and a substantial part of the radiation is absorbed along the path l_0, therefore the emitting layer in the gaseous medium is arbitrarily placed an optical distance of $\tau \approx 1$ from the observer ("Milne's criterion").

Let us now convert to a system of coordinates with its origin at the outer boundary of the solar atmosphere. This system is the most convenient for finding the emission level for the condition $\tau = 1$ along the line of sight. If

we divide equation (1) by χ, then in the new system of coordinates we have

$$\chi\, dl = -\chi\, ds,$$

so that

$$\frac{d}{d\tau} I_\lambda = I_\lambda - I_\lambda^*. \tag{4}$$

It is easy to write equation (4) as a total differential if we multiply by $e^{-\tau}$:

$$\left(\frac{d}{d\tau} I_\lambda - I_\lambda\right) e^{-\tau} = -I_\lambda^* e^{-\tau}.$$

Thus,

$$\frac{d}{d\tau}\left(I_\lambda e^{-\tau}\right) = -I_\lambda^* e^{-\tau},$$

and by integrating both sides with respect to τ over limits from 0 to τ_0 we find

$$I_{\lambda(\tau=\tau_0)} e^{-\tau_0} - I_{\lambda(\tau=0)} = -\int_0^{\tau_0} I_\lambda^* e^{-\tau}\, d\tau. \tag{5}$$

This expression can be used to calculate the specific intensity of the solar atmospheric radio emission.

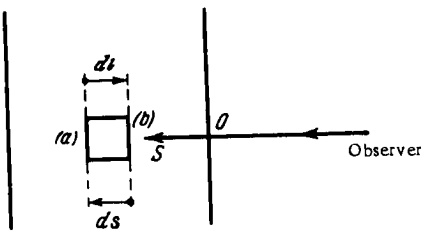


FIGURE 66

It should be noted that certain assumptions were made during the derivation of this expression. A fundamental assumption is the applicability of the Rayleigh-Jeans and Kirchhoff-Clausius formulas. These formulas hold true for media in thermodynamic equilibrium. Here, we apply them to small volumes in which it is assumed that the conditions for "local" thermodynamic equilibrium are satisfied, whereas the atmosphere as a whole cannot be assumed to be in equilibrium. The same assumption is made during calculations at optical frequencies, and it has been confirmed in part by the good agreement of the theoretical results with observations. As will be shown in the following, this also holds true for the radio range.

Some other assumptions, which are however not of fundamental significance, are as follows.

A. Equation (4) is written for a central ray. For other rays the elementary path elements ds in each layer are larger and should be replaced by $ds \sec\theta$ where θ is the angle between the observer's line of sight and a line to the center of the sun.

B. In equation (4) it is assumed that the refractive index $n = 1$; this is true for small λ (centimeter waves). Also, for small λ, $n \neq 1$ only in those

layers for which $\tau \gg 1$ (relative to the observer), but the radiation from these layers does not reach us. At longer wavelengths the refractive index must be taken into consideration.

Equation (4) without these simplifications has the form

$$\frac{d}{d\tau \cos\theta}\left(\frac{I_\lambda}{n^2}\right) = \frac{I_\lambda}{n^2} - \frac{I_\lambda^*}{n^2}. \tag{6}$$

Moreover, the refractive index is given by the expression

$$\left.\begin{aligned} n^2 &= 1 - \frac{4\pi e^2 N_e}{m\omega^2} \approx 1 - 3.18\cdot 10^9 \frac{N_e}{\omega^2} \\ \text{or} \qquad n^2 &= 1 - \frac{81\cdot 10^6 N_e}{f}, \end{aligned}\right\} \tag{7}$$

where e and m are the charge and mass of the electron in CGS units, N_e is the electron concentration per cm^3, and ω is the angular frequency ($\omega = 2\pi f$).

Since for centimeter waves $n \approx 1$, they propagate in straight lines, and the calculations according to (4) are relatively simple. On the other hand, for longer waves the path curvature complicates the picture of the radio emission in the solar atmosphere considerably. Some examples of paths for various radio wavelengths are shown in Figure 67.

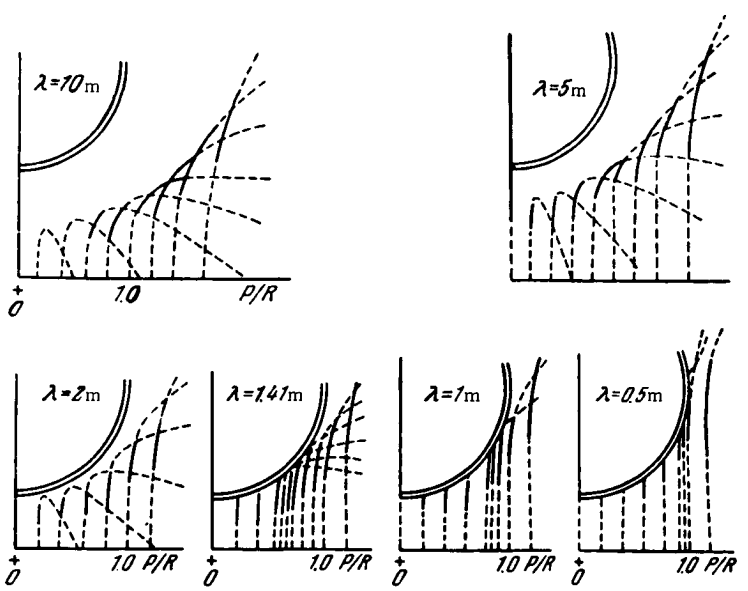


FIGURE 67

The absorption coefficient χ is calculated from the following formulas (see, for example, /18/):

$$\chi = \nu_{\text{eff}} \frac{1-n^2}{cn}, \tag{8}$$

$$\nu_{\text{eff}} \approx \frac{5.5}{T_e^{3/2}} N_e \ln\left(220 \frac{T_e}{N_e^{1/3}}\right). \tag{9}$$

Now, by combining (7), (8), and (9), we obtain

$$\chi = \xi \frac{N_e^2}{f^2 T_e^{3/2}}, \tag{10}$$

where

$$\xi = \frac{5.5e^2}{\pi cmn} \ln\left(220 \frac{T}{N^{1/3}}\right).$$

For the values of N_e, T_e, and f characteristic of the solar atmosphere, the function ξ remains fairly constant. Different authors have given values of it from 0.07 to 0.14.

In some particular cases, the transfer equation can be solved much more easily. This is so because, in the radio range, $\tau_0 \gg 1$ for the entire thickness of the solar atmosphere. This can be verified easily by substituting into (10) the values of $N_e(h)$ and $T_e(h)$ found approximately from the results of optical observations. From (5) it thus follows that for an optically deep layer

$$I_{\lambda(\tau=0)} = \int_0^{\tau_0} I_\lambda^* e^{-\tau} d\tau. \tag{11}$$

If, regardless of the preceding condition, we can be certain that the radiation reaching the layer in question from the reverse side (relative to the observer) is much weaker than the radiation emitted in the layer itself, then in (5) we will have $I_{\lambda(\tau=\tau_0)} \ll T_{\lambda(\tau=0)}$, so that expression (11) applies once again.

In order to simplify the discussion, let us analyze two idealized cases.

Case 1. If in the given atmospheric layer of the sun we can set T_e approximately constant, then I_λ^* is constant and

$$I_\lambda \approx I_\lambda^* (1 - e^{-\tau_0}). \tag{12}$$

Case 2. If, moreover, $\tau_0 \ll 1$ whereas the radiation coming in from the deeper layers is negligible, then we have

$$I_\lambda \approx I_\lambda^* \left(\tau_0 - \frac{\tau_0^2}{2} + \ldots\right) \approx I_\lambda^* \tau_0. \tag{13}$$

The second case holds true when we calculate the emission from an optically thin, hot atmospheric layer, for instance when estimating the emission of the solar corona or of coronal condensations in the centimeter range.

The thermal component of the radio emission should naturally be evaluated not only in flux units but also in units of temperature. To do this, the real emitting medium is replaced by a blackbody whose effective temperature T_{eff} is such that the specific intensity of its emission is equal to that of the real medium. Thus, after applying the Rayleigh-Jeans formula (2) and substituting expressions (11), (12), and (13) into it, we find that

$$T_{\text{eff}} = \int_0^{\tau_0} T_e e^{-\tau} d\tau \tag{14}$$

and

$$T_{\text{eff}} \approx T_e (1 - e^{-\tau_0}), \tag{15}$$

so that for $\tau_0 \ll 1$

$$T_{\text{eff}} \approx T_e \tau_0. \tag{16}$$

By taking into account (10), we can find for the last case a simple expression for T_{eff} as a function of frequency:

$$T_{\text{eff}} \sim \frac{1}{f^2} \quad \text{or} \quad T_{\text{eff}} \sim \lambda^2. \tag{17}$$

A somewhat different frequency dependence is obtained for the radiation flux, which is related to I by the equation $F = I\Omega$ (for a source with T_e constant). According to (10), (13), and (2), for an optically thin layer,

$$F \approx \text{const}. \tag{18}$$

For optically deep layers, on the other hand, $\tau_0 \gg 1$, and from (12) and (2) we have

$$T_{\text{eff}} = \text{const} \quad \text{and} \quad F \sim \frac{1}{\lambda^2}. \tag{19}$$

The close agreement between the spectral characteristics of the radiation and these relations is one of the basic proofs of its thermal origin.

The value of T_{eff} for any point on the solar surface is called the brightness temperature T_b. The mean effective temperature T_{av} of the radiation emitted by the entire sun can be found by summing T_b over the whole surface and then taking the mean.

During radio-astronomical observations it is useful to measure the effective antenna temperature. For an antenna with a small surface area the directional diagram is wide, and the antenna temperature is proportional to T_{av} and to the ratio between the solid angles of the sun and of the diagram:

$$T_A \approx T_{\text{av}} \frac{\Omega_\odot}{\Omega_A}, \tag{20}$$

where $\Omega_\odot$ is the solid angle of the sun (for radio emission), and Ω_A is the solid angle of the antenna diagram.

For very large antennas, in cases where the directional diagram is considerably smaller than the angular size of the sun,

$$T_A \approx T_b. \tag{21}$$

It is very convenient to estimate the radio flux in units of antenna temperature, since the receiver input power is related to T_A by the simple expression

$$P_{\text{rec, in.}} = kT_A\Delta f,$$

where k is the Boltzmann constant and Δf is the receiver transmission-band width.

Such an estimate is quite natural for the thermal radiation which we have been considering, but it is sometimes also applied arbitrarily to other cases. Therefore, since in reality radio sources are often not perfect blackbodies, while the radio emission may not even be thermal, considerable caution should be exercised when applying the quantity T_{eff}; this paramter should in no way be associated with the actual mechanism of the emission.

Calculations of the solar radio emission according to the above scheme have been made by various authors (see the reviews in /1/, /2/, /3/, and /9/). In these cases the vertical distribution of N_e and T_e were taken from

optical observations. The results of these calculations are plotted in Figures 68 and 69 as graphs showing the distribution of T_b over the solar surface.

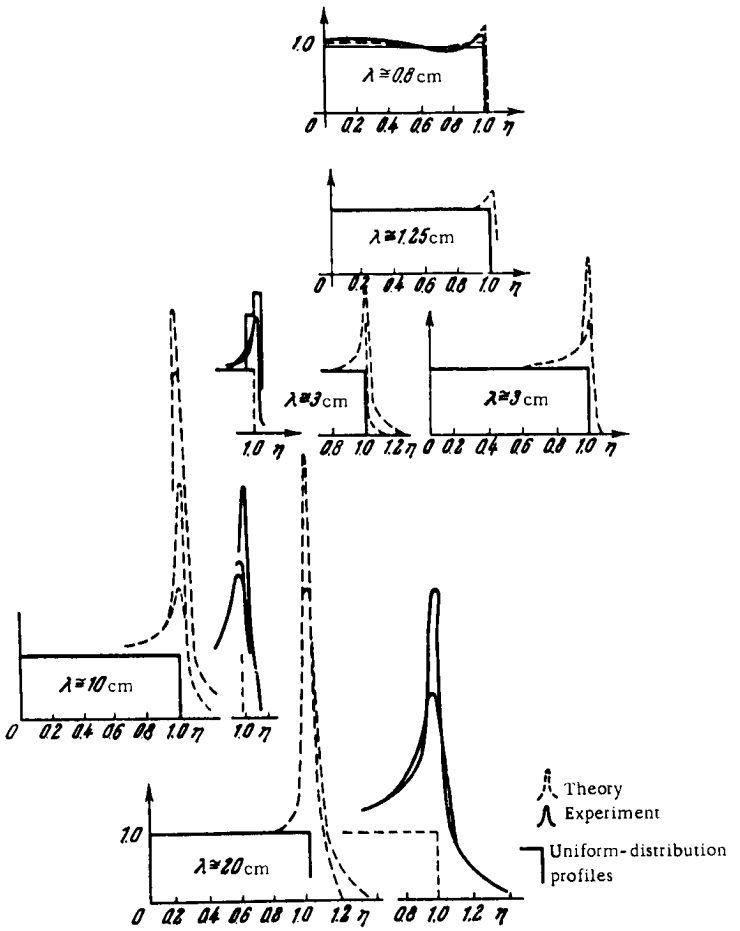


FIGURE 68

The T_b distributions shown in these figures have an interesting feature in common, namely that the "brightness" temperature increases at the limb, whereas in the optical range (especially in the ultraviolet region) a distinct limb darkening is observed. This limb-darkening effect is due to the raising of the emission level in the photosphere for noncentral rays, along with a corresponding displacement of this level into the cooler layers (positive temperature gradient). The emission level in the radio range is also higher for noncentral rays. However, the radio emission originates in layers above the photosphere, where a negative temperature gradient is maintained (the temperature T_e increases with altitude), so that correspondingly T_b at the solar limb is higher than at the center. A similar situation

is observed for the distribution of T_b in the far-ultraviolet region of the optical spectrum (X-ray emission); frequencies in this range are also emitted in the upper layers of the solar atmosphere (according to observations with artificial satellites).

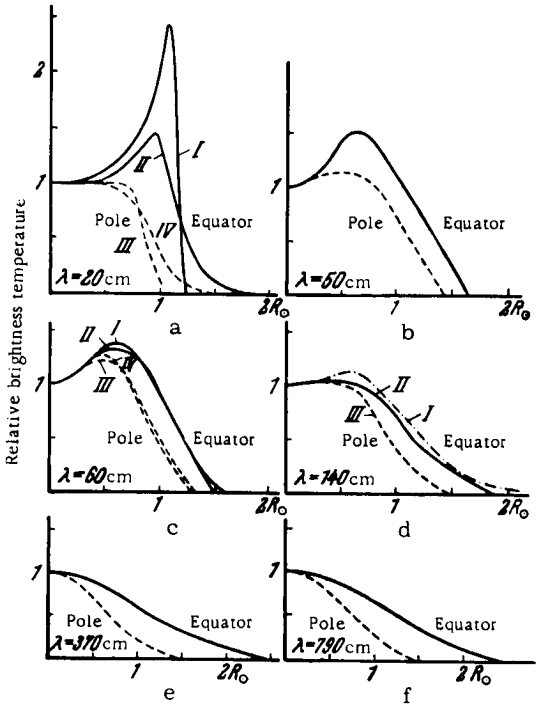


FIGURE 69

Since radio-astronomical observations specify the temperature T_e of the solar atmosphere, it is tempting to use them to construct the distribution $T_e(h)$. This task, however, is complicated by the fact that the emission levels for various wavelengths depend in turn on the distribution $N_e(h)$. Thus, since the distribution $N_e(h)$ is better known than $T_e(h)$, it is advisable to compute $T_e(h)$ from the results of radio-astronomical observations together with the optical data for $N_e(h)$.

The calculation can be made using the method of successive approximations. The distribution $N_e(h)$ is held constant, while the values of $T_e(h)$ are varied until the theoretical T_{eff} fits the observations. Such calculations have been carried out /19/ using for the distribution $N_e(h)$ in the corona ($R > 1.03\ R_\odot$) the values obtained during observations of the brightness of the coronal emission. For the chromosphere ($1.0007 < R < 1.021$) the data are less reliable, and in fact may be considered hypothetical. The calculations were based on the formulas discussed above. The experimental values of T_{av} obtained from observations and used in the calculations are given in the second column of Table 3.

As a first approximation, the temperature $T_e(h)$ is assumed to increase exponentially from 4830°K (h = 0) to 10^6 °K (h = 50,000 km). The calculated function $T_e(h)$ is plotted in Figure 70; this figure also shows the function $N_e(h)$ used in the computations. The calculations are significant in that they have been made for T_{av} (calculation) values which are in good agreement with the values obtained from a statistical processing of observations for a long period (1955 —1956). These statistical data are given in the third column of Table 3.

TABLE 3

λ, cm	T_{av} (calculation), °K	T_{av} (observation), °K
0.85	7,400	—
3.2	12,700	13,000
10	43,000	41,000
50	420,000	—

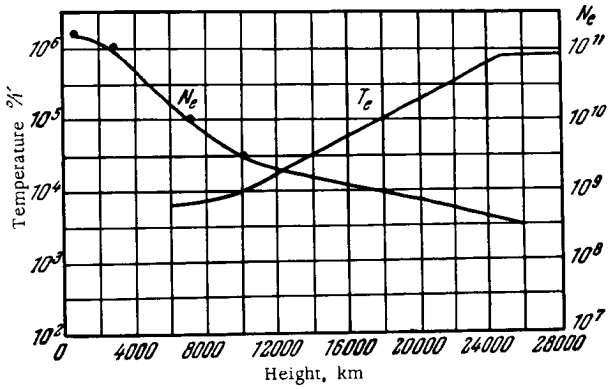


FIGURE 70

Naturally, however, a determination of $T_e(h)$ using the distribution $N_e(h)$ obtained from low-accuracy optical observations reduces considerably the worth of the independent determination of $T_e(h)$ using radio-astronomical techniques. It is therefore more meaningful to correlate the results of the radio-astronomical observations with the ratio $\frac{N_e^2}{T_e^{3/2}}$ obtained from optical observations (this approach, however, can still be criticized for the above reasons). According to optical eclipse observations, this ratio is given by the formula

$$\frac{N_e^2}{T_e^{3/2}} \approx ae^{-bh}, \tag{22}$$

where a is a coefficient. On the other hand, the absorption of radio waves is also proportional to $\frac{N_e^2}{T_e^{3/2}}$. It was found previously (14) for a central ray that

$$T_b = \int_0^{\tau_0} T_e e^{-\tau}\, d\tau,$$

where

$$\left.\begin{aligned}\tau &= \int_0^{l_0} \chi\, dl = \frac{\xi}{f^2}\int_0^{l_0} \frac{N_e^2}{T_e^{3/2}}\, dl = \frac{a}{f^2}\,,\\ a &= \xi \int_0^{l_0} \frac{N_e^2}{T_e^{3/2}}\, dl.\end{aligned}\right\} \tag{23}$$

In order to find $T_e(h)$ from these relations, $T_b(f)$ must be known. For a central ray, for which formula (14) applies, $T_b(f)$ can be found approximately from the values of T_{av} obtained during ordinary radio-astronomical observations of the mean flux emitted by the entire disk. Of course, the coronal emission must be subtracted from this quantity (it is naturally much better to find T_b according to observations with higher resolving powers, during which direct determinations of T_b at the center of the solar disk are possible).

T_b as a function of f can be expressed empirically as

$$T_b(f) = c + df^{-1}\,, \tag{24}$$

where c and d are coefficients. Next, after inserting (24) and (22) into (14), we obtain

$$cf^2 + df = \int_0^{\infty} T_e e^{-\frac{a}{f^2}}\, da. \tag{25}$$

Tables of Laplace transforms indicate that for equation (25)

$$T_e = c + \frac{d}{\sqrt{\pi a}}\,. \tag{26}$$

The substitution of (22) into (26) then gives

$$T_e = c + \frac{d}{\sqrt{\pi}}\left[\xi \int_0^{l_0} \frac{N_e^2}{T_e^{3/2}}\, dl\right]^{-\frac{1}{2}}, \tag{27}$$

and finally, by substituting (22) into (27) and seeing that $h = -l$ and $e^{bl_0} \ll 1$, we obtain the following functions:

$$\left.\begin{aligned}T_e(h) &= c + \frac{d}{a}\sqrt{\frac{b}{\pi}}\, e^{\frac{b}{2}h}\\ N_e(h) &= bT_e(h)^{\frac{3}{4}} e^{-\frac{b}{2}h}.\end{aligned}\right\} \tag{28}$$

Two sets of calculations were made using formulas (28). In one case the values of T_{av} and $\frac{N_e^2}{T_e^{3/2}}$ obtained during observations at different times /19/ were used, and in the other the values obtained simultaneously during the 1952 eclipse /20/ were used. The latter calculations are apparently more reliable.

In the first case, the coefficients in (22) and (24) had the following values: $a = 1.44 \cdot 10^{16}$, $b = 1.2 \cdot 10^{-8}$ (where the height h is in centimeters above the 1500 km level), $c = 4800$, and $d = 9.1 \cdot 10^{13}$.

In the second case:

$a = 1.45 \cdot 10^{18}$, $b = 2.19 \cdot 10^{-8}$ (for $h < 2500$ km),
$a = 3.72 \cdot 10^{17}$, $b = 1.75 \cdot 10^{-8}$ (for $h > 2500$ km),
$c = 5830$, $d = 2.84 \cdot 10^{13}$.

The coefficients in the second case were calculated using the results of both optical and radio observations during the 1952 solar eclipse. The following values of T_{av} were obtained from these:

$$T_{av} = 6\,700\ ^\circ\text{K}\ (\lambda = 0.85\ \text{cm}),$$
$$T_{av} = 12\,500\ ^\circ\text{K}\ (\lambda = 3.2\ \text{cm}),$$
$$T_{av} = 50\,000\ ^\circ\text{K}\ (\lambda = 10\ \text{cm}).$$

In both cases, in order to calculate T_b for the central ray the coronal emission T_{cor} had to be subtracted from the measured values of T_{av}. The calculated values of T_{cor} are given in Table 4.

TABLE 4

λ, cm	T_{av}, °K	T_{cor}, °K	T_b, °K
1.25	10 000	180	9 800
3.2	17 000	1 150	15 800
10	54 000	11 400	42 600

The resulting distributions $T_e(h)$ and $N_e(h)$ are given in Table 5. The data pertaining to the first set of calculations are in columns 2 and 3, and those pertaining to the second set are in columns 4 and 5.

TABLE 5

1	2	3	4	5
h, km	T_e, °K	N_e, cm^{-3}	T_e, °K	N_e, cm^{-3}
1000	5850	$2.7 \cdot 10^{11}$	—	—
2000	5900	$9.7 \cdot 10^{10}$	5041	3.10^{11}
3000	6000	$3.1 \cdot 10^{10}$	5240	$1.7 \cdot 10^{11}$
4000	6200	$1.3 \cdot 10^{10}$	5595	$9.8 \cdot 10^{10}$
5000	6750	$5.7 \cdot 10^{9}$	6250	$5.8 \cdot 10^{10}$
6000	8000	$2.7 \cdot 10^{9}$	7460	$3.8 \cdot 10^{10}$
7000	11000	—	9630	$2.43 \cdot 10^{10}$
9000	39000	—	20800	$2.3 \cdot 10^{10}$

The data in columns 2 and 3 will be useful for comparisons with mean values of $T_e(h)$ and $N_e(h)$; such mean values can be obtained according to the chromospheric models and from other complex observations which may be made in the future. However, comparisons like these should be made with caution, due to the variability of the conditions in the chromosphere.

Moreover, the spectra of the undisturbed component used in the above calculations are not entirely reliable either, because of the low accuracy of any absolute measurements of the emission flux of radio-astronomical sources. This shortcoming has not yet been eliminated, since the absolute calibration of radio telescopes is generally carried out only for individual units, and no reliable absolute standards of the relevant quantities are available /21/.

The measurements at longer wavelengths are particularly unreliable. For $\lambda < 10$ cm, we can select the most reliable of the published data (specially devised routine observations of radio fluxes and simultaneous

observations of solar eclipses at several wavelengths, which make it easy to eliminate the effects of the disturbed components) in order to derive empirical expressions for the spectrum* of the undisturbed component $F_0(\lambda)$:

for the lower boundary, approximately corresponding to the period of low solar activity ($1\text{ cm} < \lambda < 10\text{ cm}$),

$$F_{0\min} \cdot 10^{-22}\ \text{watt/m}^2 \cdot \text{cps} = 1400\,\lambda^{-1.7} e^{0.1(\lambda-1.5)}, \tag{29}$$

for the upper boundary of the spectrum, corresponding to high activity ($1\text{ cm} < \lambda < 7\text{ cm}$),

$$F_{0\min} \cdot 10^{-22}\ \text{watt/m}^2 \cdot \text{cps} = 1400\,\lambda^{-1.7} e^{0.2(\lambda-0.8)}. \tag{30}$$

Here λ is in centimeters.

A first approximation to describing the undisturbed component, assuming a spherically symmetric solar atmosphere. The above calculation constitutes a first approximation to describing the solar radio emission. The main inaccuracy here is the assumption that the different layers of the sun's atmosphere are spherically symmetric. This, as we know, is quite far from the actual situation in the solar atmosphere.

However, some of the conclusions obtained on the basis of this assumption have proved to be most valuable. First of all, the increase in T_b for the radio emission toward the solar limb has been established theoretically and verified experimentally. This increase indicates the existence of a negative temperature gradient in the solar atmosphere and corroborates the basic proposition, first advanced by Shklovskii and suggested independently by Ginzburg and by Martin in 1946, that the solar radio waves are thermal in origin. Furthermore, the high values of $T_{av} \approx 10^6$ °K determined from radio observations in the meter range confirm the existence of very high electron temperatures in the corona, which hitherto could only be inferred rather uncertainly from various optical observations.

However, from radio observations no detailed conclusions have been obtained concerning the distribution of the solar-atmosphere parameters with the height h. The main difficulty lies in the somewhat arbitrary choice of the distributions $N_e(h)$, and particularly $T_e(h)$, in the solar atmosphere, due to the difficulty involved in making observations in the optical range. This especially applies to the chromosphere. The evaluation of the results of radio observations often reduces just to concluding that they are consistent with the models proposed. Such a conclusion (which by no means guarantees that the inferences made from the observations are unambiguous) is nevertheless frequently offered as proof of the validity of some particular model of a stratified chromosphere. The following conclusions drawn from observations can be regarded as a sufficiently reliable first approximation, assuming a stratified and spherically symmetric solar atmosphere.

A. The radio-emission level for any particular wavelength becomes further from the photosphere the higher the wavelength .** That this is

* Since the solar emission is the resultant of several random emission processes, therefore we have here, strictly speaking, not a spectrum with monochromatic components but rather a variation of the spectral density of radiation with wavelength. We will use the term "spectrum" in the following discussion, but always this reservation will be implied.

** Strictly speaking, a frequency band about some mean wavelength is involved here.

the case follows from all the theoretical schemes available and also from observations. An example of the distribution of radio-emission levels and the contribution of each level are shown in Figure 71. From the formulas derived above it is easy to show that even for arbitrary variations in the distributions $N_e(h)$ and $T_e(h)$, over a range bounded by physically unacceptable values, the emission levels still follow a pattern in which virtually the entire radiation in the 2 to 10 cm range originates in the upper chromospheric layers (5000 km $< h <$ 13,000 km), whereas all the radiation with $\lambda < 2$ cm comes from the lower chromosphere ($h <$ 7000 km). Wavelengths of $\lambda \geqslant 50$ cm originate almost entirely in the solar corona: $h > (1.8 \text{ to } 2.0) \cdot 10^3$ km.

The observation results confirm these conclusions. Especially instructive is a determination of the sun's "radio diameter." The amount by which the radio diameter exceeds the optical diameter of the sun indicates the upper limit for the emission level, since here we are actually recording the radiation emitted along a tangent to the sun, that is, along a path with a maximum value of τ. Some radio values of the solar radii obtained during eclipse observations are listed in Table 6, where R_r is the solar radius found from radio-astronomical observations, $R_\odot$ is the radius of the photosphere, and h is the height of the outer boundary of the emitting layers.

TABLE 6

λ, cm	Time of observation	$R_r/R_\odot$	h, km
0.8	30 June 1954	1.0085 ± 0.0025	~ 6000
0.86	30 June 1954	1.007 ± 0.0005	~ 5000
3.2	25 Feb. 1952	1.029	~ 20 000
3.2	1 Sept. 1951	1.030 ± 0.003	~ 21 000
3.2	30 June 1954	1.032 ± 0.005	
3.2	31 July 1962	1.034 ± 0.003	
		1.033 ± 0.003	~ 23 000
4.5	31 July 1962	1.034 ± 0.005	
		1.032 ± 0.005	
10	30 June 1954	1.035 ± 0.008	~ 25 000
50	30 June 1954	1.06 ± 0.01	~ 42 000
150	20 May 1947	~ 1.35	~ 250 000

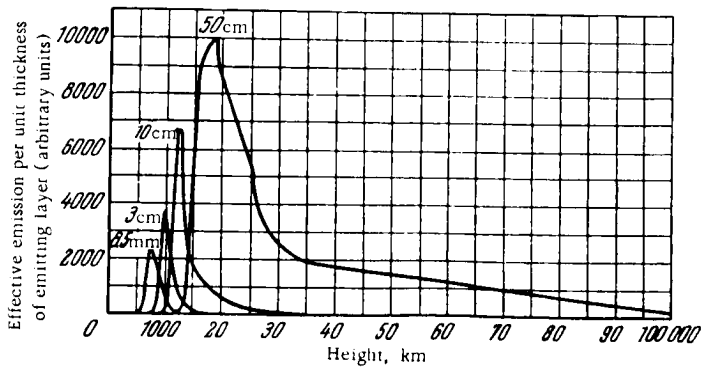


FIGURE 71

No less illustrative are graphs showing the variation of the ratio K of the residual fluxes F_{res} at λ = 10 cm and λ = 3.2 cm, for various phases of

a solar eclipse (Figure 72). The increase in K is definitely greater than any possible measurement error, and it indicates a rapid growth in the contribution of radiation at $\lambda = 10$ cm for $\eta = R_{\mathbb{C}}/R_{\odot} > 1.025$ ($h > 17{,}000$ km).

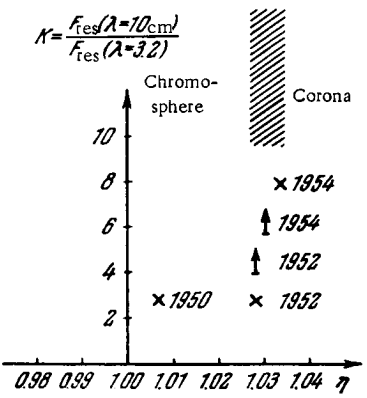


FIGURE 72

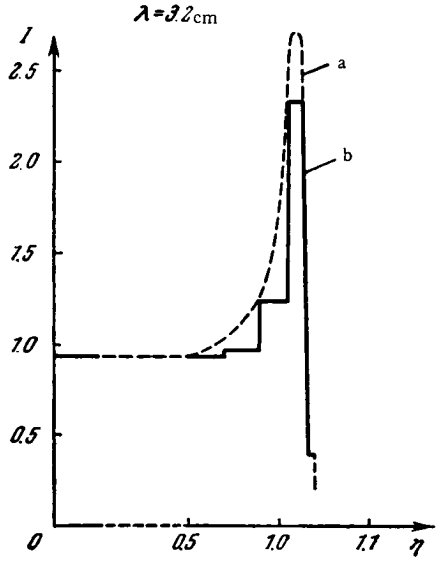


FIGURE 73

The fact that the undisturbed component at $\lambda < 10$ cm does not depend much on the state of the highly variable solar corona enables us to compare the residual radio fluxes obtained during the eclipses of different years. The ratios formed by the differences in residual flux divided by the differences in the geometrical areas of the sun covered by the moon during different eclipses give relative values of T_b for the corresponding disk zones. This type of comparison gave the distribution of T_b over the solar disk ($\lambda = 3.2$ cm) shown in Figure 73 (curve b); it is in good agreement with the previously described calculation results (curve a). (This distribution is still classified within the framework of the first approximation, as defined above.)*

B. The electron temperature T_e of the lower chromosphere is actually comparatively low, not exceeding 10,000°K. On the basis of the preceding discussion, this conclusion follows from the observation results listed in Tables 6 and 7 (the data include the increased emission at the solar limb).

This result concerning the temperature of the lower chromosphere is very important, since it confirms to a considerable degree the results of certain optical observations indicating a low chromospheric temperature, contrary to the opinion currently held by some authorities that temperatures as high as $T_e = 30{,}000$°K are possible in the layer $h \approx 1500$ km.

* Recent results show that the distribution of T_b over the solar disk varies somewhat with the solar activity, even at $\lambda = 3.2$ cm. Thus, if we compare F_{res} only for eclipses of the "quiet" sun, then we obtain a much smaller increment of T_b (no more than 20-30%) at the limb than the one shown in Figure 73. (Note added during proofreading.)

C. The electron temperature of the corona is high, and the temperature gradient in the outer coronal layers is positive. The high value of T_e in the inner corona is responsible for the fact that the increase in T_b at the solar limb, even in the centimeter and decimeter ranges, is substantially greater than the temperature of the respective emission layers. Observations at $\lambda > 50$ cm give $T_{av} > 5 \cdot 10^6$ deg.

TABLE 7

λ, cm	T_{av}, °K	λ, cm	T_{av}, °K
0.18	5300	0.43	7000 ± 700
0.215	5430	0.86	8500
0.273	5500	0.86	6740
0.3	5900	0.86	5300
0.4	8000	1.25	10 000 — 11 000
0.43	9600		

The distribution of T_b was calculated using a model of an isothermal corona and assuming the temperature in all the coronal layers to have a constant value of about 10^6 deg. This model leads to the conclusion that the brightness maximum of the solar disk is located right at the limb itself or even somewhat outside the boundary of the photosphere. This is quite obvious from Figure 74. T_b starts diminishing only for rays which are so far from the photospheric boundary (away from the disk) that the solar atmosphere becomes transparent to them. This very significant conclusion related to the assumption of an isothermal corona does not agree with observations, as has been emphasized in /22/. Actually, the T_b maximum in the decimeter range is displaced toward the center of the disk. The corresponding data are given in Table 8.

TABLE 8

λ, cm	$\frac{R_{T_b=\max}}{R_\odot}$	$\frac{T_{b\max}}{T_{b\,center}}$	Year of observation
21	0.96	1.6	1953—1954
24	1.0	2.0	1952
24	1.0	2.2	1956
50	0.65	1.6	1955—1956
60	0.80	1.5	1954—1955
60	0.60	1.5	1954
140	—	1.0	1952
145	0.65	1.1	1953—1955

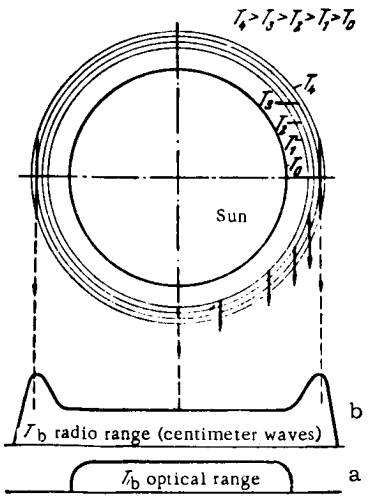


FIGURE 74

Such a displacement seems quite natural if we assume a positive temperature gradient in the outer coronal layers. In this case, as the emission-level height is increased, the radiation passes successively (Figure 75) through a low-temperature layer (T_b), then through a hot layer ($T_2 > T_1$), and then through another cooler layer ($T_3 < T_2$). Roughly speaking, for the shortest waves, the T_b gradient is negative throughout the emitting domain, and the brightness maximum of the disk is situated outside of the photosphere; for the longest waves, the temperature gradient in all the emitting layers is negative and limb darkening is observed; in the intermediate layers both of these effects take place, and their combined action apparently displaces the T_b maximum toward the center of the disk.

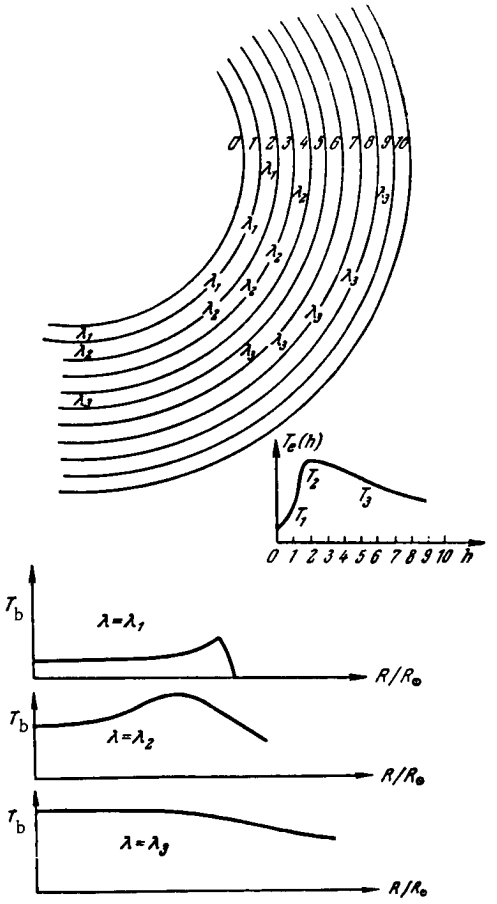


FIGURE 75

A second approximation to describing the undisturbed component, taking into account the ellipticity of the solar corona. All the preceding data, based on the assumption of a stratified, spherically symmetric solar atmosphere, are very interesting,

even though they remain within the framework of a first approximation. Optical observations show, however, that the solar corona is not spherically symmetric, but rather is characterized by a complex structure.

For example, observations of various coronal elements in 1914 and 1941 gave electron concentrations which differed by factors of 2 or 3 from one another; furthermore, the electron concentrations in coronal rays may be as much as ten times greater than the mean value at distances up to $4R_\odot$ from the photosphere.

It is also a known fact that in years of minimum solar activity the concentration of electrons near the equator N_{eeq} is considerably higher than the polar concentration N_{ep}. The ratio P of these concentrations for the inner corona may be as high as 1.3 to 2.2. It is lower only for a few months near the epoch of maximum solar activity. Some values of N_e obtained from optical observations for the equator and the poles are given in Table 9.

TABLE 9

$\eta = \frac{R}{R_\odot}$	Minimum solar activity			1941 — 1957			Enhanced solar activity		
	van de Hulst*			Waldmeier **			Newkirk †		
	$N_{ep} \cdot 10^{-6}$, cm^{-3}	$N_{eeq}{}^{10-6}$, cm^{-3}	P	$N_{ep} \cdot 10^{-6}$, cm^{-3}	$N_{eeq}{}^{10-6}$, cm^{-3}	P	$N_{ep} \cdot 10^{-6}$, cm^{-3}	$N_{eeq}{}^{10-6}$, cm^{-3}	P
1.00	174	227	1.3	—	—	—	240	530	2.2
1.03	127	178	1.4	106	255	2.4	—	—	—
1.06	87.2	132	1.6	39	96	2.4	—	—	—
1.10	53.2	90.0	1.7	29.4	72.8	2.4	—	—	—
1.125	—	—	—	—	—	—	77	170	2.2
1.15	—	—	—	20.0	53.0	2.6	—	—	—
1.20	16.3	39.8	2.4	13.2	35.9	2.7	—	—	—
1.25	—	—	—	—	—	—	32	70	2.2
1.3	5.98	21.2	3.5	7.1	23.3	3.3	—	—	—
1.375	—	—	—	—	—	—	15	34	2.25
1.4	—	—	—	4.7	15.5	3.3	—	—	—
1.5	1.41	8.3	6.0	—	—	—	8.6	19	2.2
1.6	—	—	—	2.6	8.1	3.1	—	—	—
1.625	—	—	—	—	—	—	5.4	12	2.2
1.7	0.542	4.0	7.4	—	—	—	—	—	—
1.750	—	—	—	—	—	—	3.5	7.7	2.2
1.8	—	—	—	1.7	4.9	2.9	—	—	—
1.875	—	—	—	—	—	—	2.4	5.2	2.15
2.0	0.196	1.58	8.0	1.4	2.9	2.0	1.6	3.6	2.2
2.6	0.040	0.374	9.3	—	—	—	—	—	—
3.0	0.017	0.176	10.3	—	—	—	—	—	—
4.0	0.004	0.050	12.5	—	—	—	—	—	—

* Coronal-brightness observations during eclipses. The ratio P is apparently somewhat too low.

** Red-line observations, 6374 Å.

† Extra-eclipse observations of the corona (end 1956 - 1957) using a K-coronameter, λ = 5300 Å.

These features of the coronal structure will definitely have some influence on the solar radio emission. The effect will be the strongest in the decimeter and meter ranges, since this radiation originates in the corona; the effect on the centimeter range will be less pronounced, in proportion to the contribution of the coronal radiation at these wavelengths.

Let us first consider the influence of coronal inhomogeneities on the results obtained according to the first approximation described above. The presence of an inhomogeneity with increased N_e naturally raises the optical depth τ to a level higher than the values characteristic of a homogeneous atmosphere, the same distances being traversed by a ray. Because of this increment, the emission levels for a given wavelength are displaced

into higher layers of the solar atmosphere. In order for the previous distribution curves to apply, we must assume formally that the presence of inhomogeneities shifts all the preceding results into a range of higher wavelengths, namely that the distributions originally obtained for meter waves actually correspond (after allowing for the inhomogeneities) to decimeter waves.

Inhomogeneities were first taken into account when it was established from observations that the observed decrease of T_b toward the solar limb (at higher wavelengths) was more gradual than that predicted by theoretical calculations. This effect, and also the apparent sizes of the radio disks on meter and decimeter waves, were found to be much greater than that to be expected from the theory based on van de Hulst's distribution of N_e in the corona.

Finally, a direct proof of the existence of deviations from this distribution was obtained from observations of the occulation of the radio Crab Nebula (Taurus A) by the solar corona (observations of Vitkevich and Hewish). According to these observations, electron inhomogeneities exist in the corona between $R = 4.5R_\odot$ and $R = 20R_\odot$. The calculations for meter waves, but in which the data obtained for the supercorona $R > 5R_\odot$ were extrapolated to the region $R < 5R_\odot$, also gave results which agreed well with the observations.

The coronal inhomogeneities may be related to the solar activity, and if this is the case the radio emission must be regarded as a variable component of the emission. Such a relation to the activity has been definitely established for inhomogeneities arising in the lower corona in regions above sunspots, and it will be discussed in §58.

The second feature of coronal structure, namely the increased mean concentration of electrons near the equator in comparison with the polar concentration, is responsible for the asymmetry of the solar radio emission (its ellipticity), an effect which is especially pronounced at higher wavelengths.

The ellipticity of the sun was clearly observed during one of the first observations of a solar eclipse, at two points in the range between $\lambda = 50$ cm and $\lambda = 1.78$ m. Detailed results were later obtained for shorter wavelengths as well. The results of some observations which evaluate the ellipticity in terms of the ratio of the ellipse axes, the radio emission of the sun being thereby approximated with the assumption that T_b is constant, are given in Table 10. The distributions of T_b near the equator and at the pole for 9.4 cm and 20 cm waves are also illustrated in Figures 76 and 77.

Ellipticity in the distribution of $N_e(h)$ is also observed in the supercorona, at distances up to $17R_\odot$ from the photosphere; it has been shown to vary with the solar activity.

Consequently, the results of radio-astronomical observations corroborate the optical data on the shape of the solar corona.

A third approximation to describing the undisturbed component, taking into account the filamentary structure of the solar chromosphere. A third approximation is achieved when we realize that the chromosphere is not a spherically symmetric stratified formation; actually, it consists of filaments, mostly radial, which differ in their physical properties.

The "jagged" appearance of the chromosphere was observed optically a long time ago. After this, chromospheric models were repeatedly proposed, but though they were empirically and theoretically sound, still they differed greatly from one another. The most agreement was reached in the estimates of the filament length (4000 to 7000 km) and thickness (1000 to 1500 km). On the other hand, the variety of electron concentrations and temperatures obtained from optical observations and from calculations can be satisfied by almost any of the radio-astronomical observations available.

TABLE 10

λ, cm	Axial ratio of ellipse	Method of observation	Date	Remarks
3.2	<1.1	Measurement of second harmonic, with antenna diagram revolving along solar perimeter during an annular eclipse	19 IV 1958	High solar activity
3.2	1.12	Same method, no eclipse	1961	Low solar activity
9.4	See Figure 76	Eclipse observation	20 VI 1955	
20	$\frac{1}{1.33}$	Eclipse observation at two points	30 VI 1954	
21	$\frac{1}{1.25}$	Interferometric observation	1953	
50	$\frac{1}{1.3}$	Eclipse observation at three points	1 XI 1948	
50	$\frac{1}{1.25-1.30}$	Interferometric observations	1953—1954	
60	$\frac{1}{1.25}$	Interferometric observations	1953—1954	
150	$\frac{1}{1.25-0.04}$	Eclipse observation	30 VI 1954	
178	$\frac{1}{1.5-1.6}$	Eclipse observation at two points	1 XI 1948	

The electron temperature in the filaments (spicules) has been given alternatively as 10,000°K and 58,000°K. In some cases the interspicular space is assumed to be filled with hotter material than the material of the spicules themselves, while in other cases it is assumed to be cooler. These differences are apparently explained by the fact that the electron temperature is obtained from optical observations, and that always various more or less arbitrary assumptions are made (constancy of temperature or equality of the eddy velocities for filaments with different temperatures, etc.). In such cases the calculated values of T_e are naturally dependent on the assumptions made, and so it is only with great caution that they can be compared with radio observations. Radio-astronomical methods, on the other hand, have the distinct advantage that they permit direct measurements of the effective antenna temperature, which is closely related to the electron temperature T_e. Unfortunately, in such measurements the height of the emission layer cannot be determined with much accuracy.

As observed previously, the chromospheric values of T_e are estimated using the effective radiation temperatures for $\lambda < 3.2$ cm. These temperatures are fairly low, not exceeding 10,000°K for $h <$ (3 to 6)$\cdot 10^3$ km. Such values, however, represent averages for the entire disk. The complex structure of the chromosphere may result in a quite unusual distribution of T_b over the solar disk, and this will be amenable to observation only when high-resolution instruments are used.

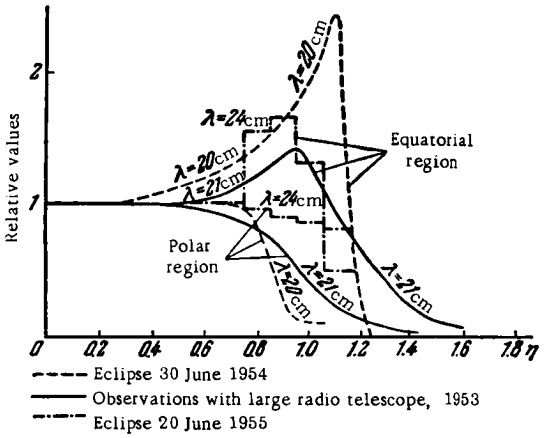


FIGURE 76

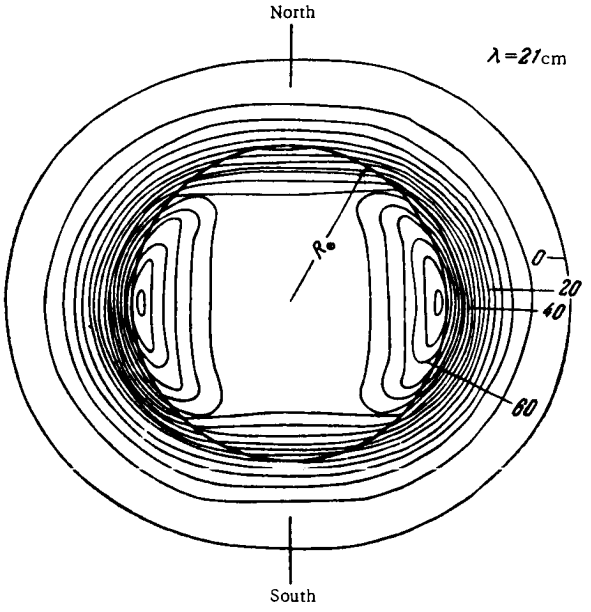


FIGURE 77

Let us now turn to Figure 78, which shows a schematic diagram of chromospheric filaments all having the same height and all directed radially.

The gas temperature T_2 in the filaments is substantially lower than the temperature T_1 of the gas in the space between them. The electron concentration N_2 in the filaments is much higher than N_1. These differences in the physical properties of the two media lead to an appreciable difference in the optical thicknesses characterizing their respective emission levels (see §55); here $\tau_1 \ll \tau_2$.

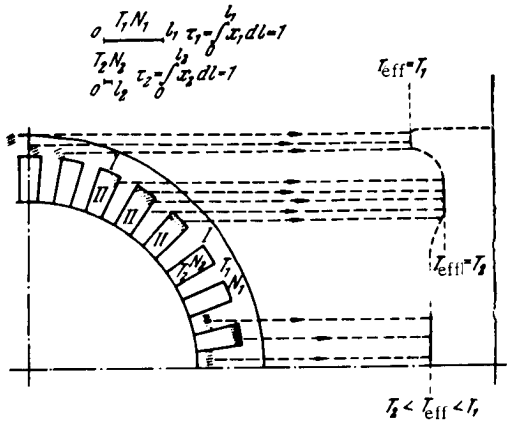


FIGURE 78

In Figure 78 a bounded region I has been designated for the sake of simplicity. According to this representation, the radiation reaching us from the center of the solar disk originates partly in region I and partly in region II (cross-hatched areas). The effective temperature of the disk center will thus have some value between T_1 and T_2, depending on the density of the filaments in the chromosphere.

Here it is assumed that the gradient at the boundary of a spicule is small, and that the radio waves are almost entirely absorbed before being reflected from the denser medium. In other words, the spicules are assumed to be "blackbodies" in the relevant frequency range. The effective temperature of the radiation emitted near the limb will be determined essentially just by the temperature of the filaments. The radiation from Region I has little effect in this case, since the deep gas layers at a temperature T_1, for which $\tau_1 > 1$, are screened by the "cold" filaments. Finally, at the limb itself, the radiation is emitted by region I and is not intercepted by any of the filaments. Thus, in this region, just as for the spherically symmetric atmosphere, the temperature increases and forms a "bright limb" around the solar disk.

Such a distribution has been confirmed by observations during an eclipse, for $\lambda = 0.8$ cm, and this in fact led astronomers to take into account the filamentary structure of the chromosphere in their interpretation of the radio observations. Once they have been confirmed, radio-astronomical data can also be used to substantiate the results of optical observations, and this has led to the idea of a structurally complex chromosphere with cold filaments.*

* The low dependence of the radio radius on wavelength in a certain part of the centimeter range (see Table 6) may possibly be caused by the influence of the propagation boundary for cold and dense opaque filaments.

In the $\lambda = 0.43$ cm range no increase in the radio brightness was observed at the center of the disk. This result, which has a measurement error of less than 1%, can be explained by the fact that the two media, both the first and the second, become so transparent that the main part of the radiation for all ray directions is emitted near the photosphere, that is, in a layer whose temperature is low at all points of the disk.

The lack of central brightening for longer waves can be explained by the opposite phenomenon, namely that the absorption in medium I is quite strong, so that the levels with $\tau \approx 1$ are too high for the filaments to have any significant influence.

Radio observations at $\lambda = 0.43$ cm and $\lambda = 0.86$ cm have led to a model of a two-component chromosphere with cold spicules and hot interstitial gas. The temperature of the spicules in this model was taken as 6400°K, while $N_e = 8.91 \cdot 10^{10}\ \mathrm{cm}^{-3}$ near the base and $2.21 \cdot 10^{10}\ \mathrm{cm}^{-3}$ for $h > 4000$ km. The gas filling the interspicular spaces in this model has $T_e = 6400$°K for $h = 2000$ km, with $T_e = 19{,}000$°K for $2000 < h < 4000$ km, and $19{,}000\text{°K} < T_e < 10^6$ °K for $4000\ \mathrm{km} < h < 10{,}000$ km. With these parameters, the distribution of radio emission at $\lambda = 0.43$ cm and $\lambda = 0.86$ cm over the solar disk is in satisfactory agreement with the results of the previously mentioned observations.

It should be noted here that dense filaments (or spicules) whose refractive indexes $n = 0$ need not be "blackbodies," since, on the contrary, they may reflect a considerable part of the radio emission. A similar phenomenon is possible, despite the strong absorption of centimeter waves in an ionized gas, provided the gradient of N_e at the boundaries of the spicules is sufficiently high. The disk area in which the spicules are located near the limb may then reflect, so that the radiation received from such areas will have a temperature depending on the contribution of the overlying transparent layers only. Since this contribution is very small for short centimeter waves (it is only about a few hundred degrees), the spicules will play a cooling role in this range, not due to their low temperatures but due to their reflecting properties. This approach explains the decrease of T_b between the center and the limb in the same way as previously done in Figure 78, the only difference here being that T_b is not the spicular temperature but rather the effective temperature of the radiation emitted along the path of a ray in the overlying layers. In this case, the ray paths in Figure 78 should not terminate in spicules, but are reflected from them like the rays of longer λ, as was shown in Figure 61.

The possibility of interpreting the properties of the spicules in different ways makes it difficult to formulate an unambiguous relation between the radio-astronomical and optical data. For example, the multicomponent chromospheric model based on optical observations includes four different kinds of filaments:

1) metallic with $N_e < 10^{9}\ \mathrm{cm}^{-3}$ and $T_e \approx 6000$°K;
2) hydrogen with $N_H \approx 10^{13}\ \mathrm{cm}^{-3}$ and $T_e \approx 8500$°K;
3) helium with $N_e \approx 10^{10}\ \mathrm{cm}^{-3}$ and $T_e \approx 25{,}000$°K;
4) subcoronal with $N_e \approx 10^{10}\ \mathrm{cm}^{-3}$ and $T_e \approx 150{,}000$°K.

It is easy to calculate from formulas (3) and (10) that for $\lambda < 3.2$ cm only the helium and hydrogen filaments (if in the latter we assume $N_e \approx 10^{12}$ cm) are opaque. In this case, however, the mean total radiation is too hot, and to fit the results of the radio observations ($T_b \approx (7$ to $8) \cdot 10^3$ °K) we must assume that the material between the filaments has $T_e < 6000$ °K instead of

the higher temperature according to this model. Thus, in order to satisfy the observed values of T_b, we must assume that the dense spicules are hotter than the gas between them, that is, that central brightening of the disk is impossible. This difficulty can be eliminated, for example, by assuming that the dense filaments reflect; then, according to (7), it is sufficient to take $N_e \geqslant 4 \cdot 10^{12}\,\mathrm{cm}^{-3}$ (for $\lambda \approx 0.8\,\mathrm{cm}$).

The reader should be warned at this point that the above radio data obtained during solar eclipses give results for the emission distribution over the disk which are not reliable enough. These data therefore cannot be regarded as conclusive, and the spread of the parameter values for the filamentary chromosphere must be narrowed as a result of consistent and simultaneous observations in both the optical and radio ranges, using high-resolution radio telescopes.

§58. **The slowly varying disturbed component of the solar radio emission.** Observations show that the contribution which the slowly varying component makes to the general radio flux of the sun depends strongly on the solar activity; it is particularly closely related to sunspot formation. The appearance of a spot group which is quite large in area is, as a rule, accompanied by an appreciable enhancement of the radio emission from a local region located approximately over the group. This enhancement may be by a factor of ten, or even a hundred, and it has its maximum in the meter and decimeter ranges. This unusual radio contrast differs from the development of sunspot activity according to its characteristics in the optical range, where spot regions are cooler than the photosphere and where the temperature difference between spots and photosphere is never greater than 30 to 40%. This component becomes less noticeable in the centimeter range and is difficult to detect for $\lambda < 1\,\mathrm{cm}$.

Consequently, in the radio range, at moderately high wavelengths, the solar radiation flux will vary much more than in the optical region of the spectrum. Radio-astronomical methods are therefore better adapted to the study of certain solar phenomena than optical observations are.

The principal characteristics of radio sources above sunspots. Radio-astronomical observations on centimeter and decimeter waves, even though made using small antennas, have been carried out for a long enough period that a statistical processing of the results is possible. In addition, sporadic observations during solar eclipses have established the following properties of the radiation emitted by local radio sources over spot groups, which apparently constitutes the main part of the slowly varying component of the solar radio emission: 1) on the average, the emission from sources above spot groups varies slowly in time, often remaining fairly constant for a few days; 2) the radio flux emitted by such sources (called "*A* sources" in the following) is closely related to the spot area of the corresponding group (the source is generally located directly "above" this group); the greater the total spot area, the greater will be the radio flux; 3) the dimensions of *A* sources in the centimeter range are approximately equal to the dimensions of the spot groups, whereas in the decimeter range the sources are comparable with the facular fields around the spot groups; 4) the effective radiation temperature of an *A* source in the centimeter range (for very short wavelengths) and decimeter range is comparable with the temperature of the solar corona; 5) part of the radiation emitted by these sources (up to 50% in the centimeter range) is circularly

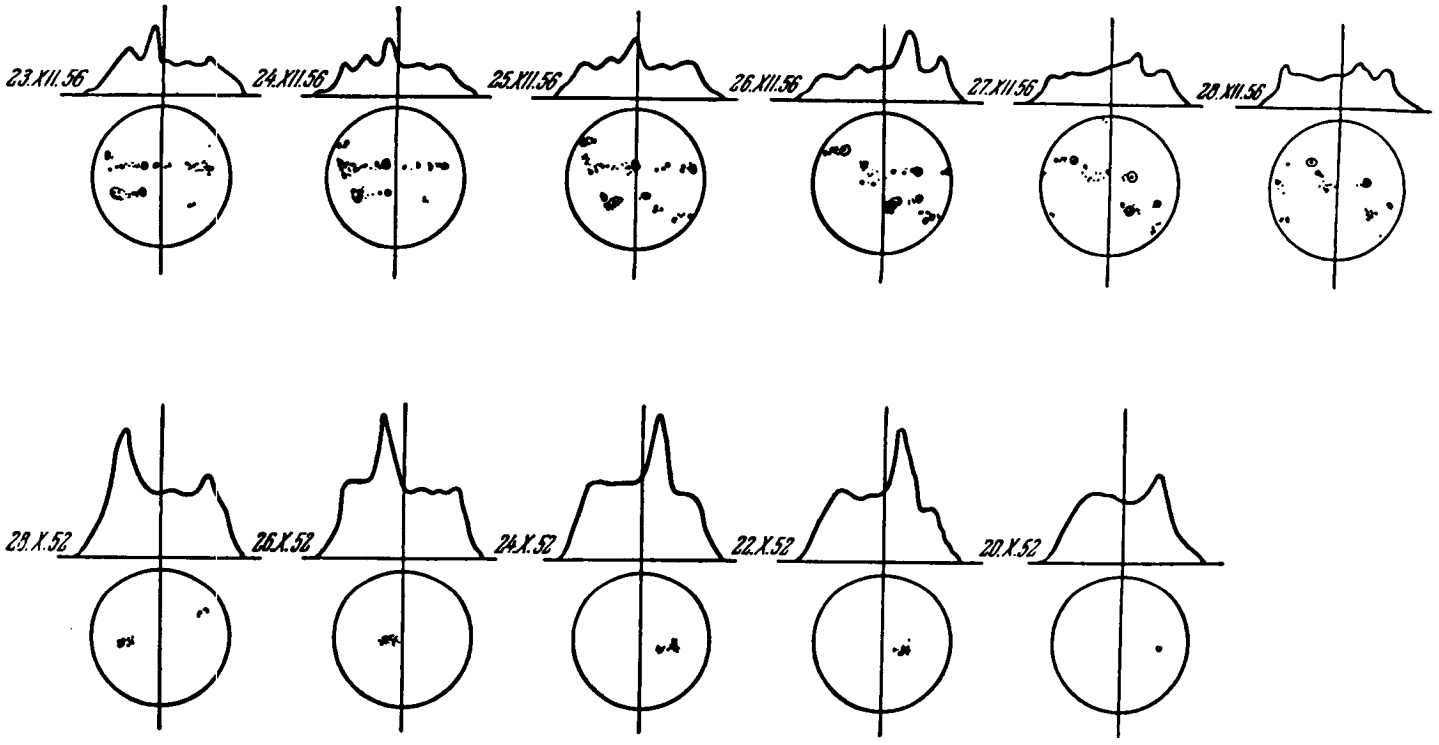

23.XII.56
24.XII.56
25.XII.56
26.XII.56
27.XII.56
28.XII.56
29.X.52
26.X.52
24.X.52
22.X.52
20.X.52

FIGURE 79

polarized; 6) the radiation flux at moderately long wavelengths depends considerably on the angle between the observer's line of sight and the perpendicular to the spot group (to the surface of the photosphere); this flux decreases to 1/3 or 1/4 as a source moves from the center to the limb; 7) the spectrum of the radiation from an *A* source has a maximum in the region $4\,\text{cm} > \lambda > 8\,\text{cm}$; 8) the radio sources rise above the limb before the corresponding spot groups do.

Recent observations with large high-resolution radio telescopes have not only confirmed these general characteristics but also have enabled a more accurate description of the corresponding properties. These observations also have supplied new data, which were unobtainable by previous observers; in particular, they established the source height h, they verified the constancy of the radiation fluxes from individual sources (as distinct from the overall emission for several sources), they specified the distribution of the brightness temperature T_b over the solar disk, etc.

The main findings of these observations so far are as follows: 1) the emission of an individual source is very stable in time, often remaining almost constant for a few days. The position of an *A* source corresponds to the position of the optically observable spot group. Numerous observations have without exception indicated such a correspondence. It is also obvious from the examples in Figures 79 and 80,* in which the black regions represent sunspots; 2) the size of an *A* source in the centimeter range corresponds approximately to the size of the spot group, and it generally does not exceed 1 to 3 minutes of arc. The sources emitting circularly polarized radiation are even smaller; 3) the height of an *A* source (the center of gravity of the emission) for moderately long waves is equal to 0.05 to 0.07 of the photospheric radius; 4) the radio sources rise above the limb before the optically visible spot groups do, and the lead time increases for higher wavelengths. In some cases the lead time has been found to be as much as three days at $\lambda = 21$ cm, as much as a day and a half to two days at $\lambda \approx 10$ cm, and as much as a day to a day and a half at $\lambda \approx 3$ cm.

We will now describe in more detail these characteristics and also the methods for determining them.

1) The constancy of the radiation flux emitted by the sources and the relation of the flux to the sunspot area S_p (in fractions of the solar hemisphere) were found by means of systematic measurements of the total radio flux of the sun F_Σ. A subsequent statistical comparison of F_Σ with the optically observed total sunspot area ΣS_p, for a long period, showed high correlation between the variations of ΣS_p and F_Σ. At $\lambda = 3.2$ cm, for example, the correlation coefficient r was found to be about 0.79 (November 1948 to March 1949), and at $\lambda = 10$ cm it was found to be $r \approx 0.78$.

An even better correlation with the solar activity was observed for the shift of the center of gravity of the solar radio emission relative to the geometrical center of the sun. Local radio sources situated a certain distance θ_s from the geometrical center of the sun emit an additional radiation flux F_s and thus distort the symmetrical distribution of the overall radiation. The center of gravity for the emission of the disturbed sun is

* Figure 79, top: Pulkovo observations, $\lambda = 3.2$ cm; Figure 79, bottom: observations at Sydney, $\lambda = 21$ cm. Figure 80, top left: T_b distribution at $\lambda = 21$ cm; Figure 80, top right: chart of sunspots and chromospheric fields; Figure 80, bottom: "isophotes" of magnetic-field strength, with dashed lines for S and solid lines for N. The dash-dot lines in the top right and bottom charts of Figure 80 retrace the outline of the solid line in the top left chart.

displaced toward the strongest sources by an amount $\Delta\theta$, which can be as high as 4 or 5 minutes of arc in the range $3\,\text{cm} < \lambda < 50\,\text{cm}$. The relation between $\Delta\theta$, the radiation fluxes, and the position of the sources is determined using the formulas for calculating the center of gravity of plane figures:

$$\Delta\theta = \sum_{i=1}^{n} F_{si}\theta_{si} \Big/ (F_0 + \sum_{i=1}^{n} F_{si}) = \sum_{i=1}^{n} \theta_{si}\frac{F_{si}}{F_0} \Big/ \left(1 + \sum_{i=1}^{n} \frac{F_{si}}{F_0}\right), \qquad (31)$$

where F_0 is the flux emitted by the quiet sun.

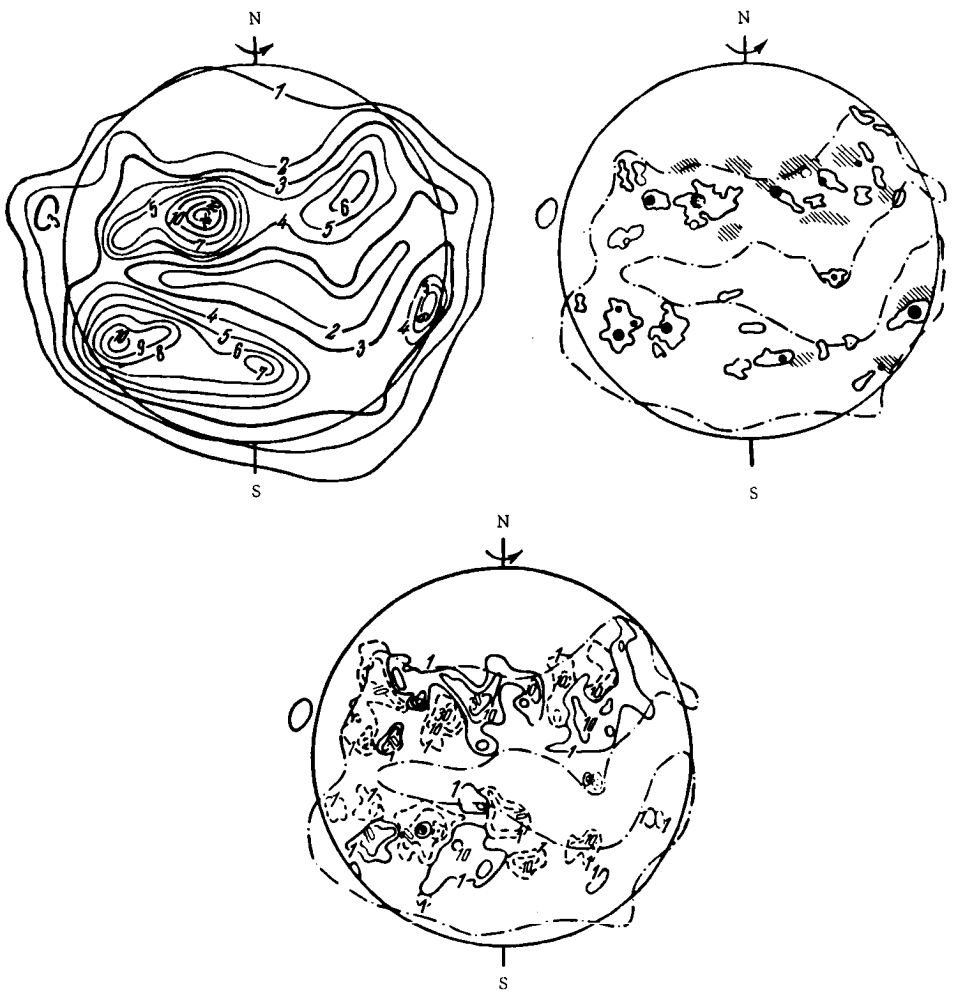


FIGURE 80

Both the fluxes from the radio sources and the distances of the sources from the center of the sun enter into expression (31). Therefore, the close correlation between $\Delta\theta$ and $S_p\theta_s$ implies that there is also a high correlation

not only between F_s and S_p but also between the positions of the radio sources and the positions of the spot groups whose values of S_p have been used. Actually, the coefficient for the correlation between $\Delta\theta$ and $S_p\theta_s$, calculated for the period from 4 December 1956 to April 1957, was found to be $r \approx 0.77$ (at $\lambda = 3.2$ cm).

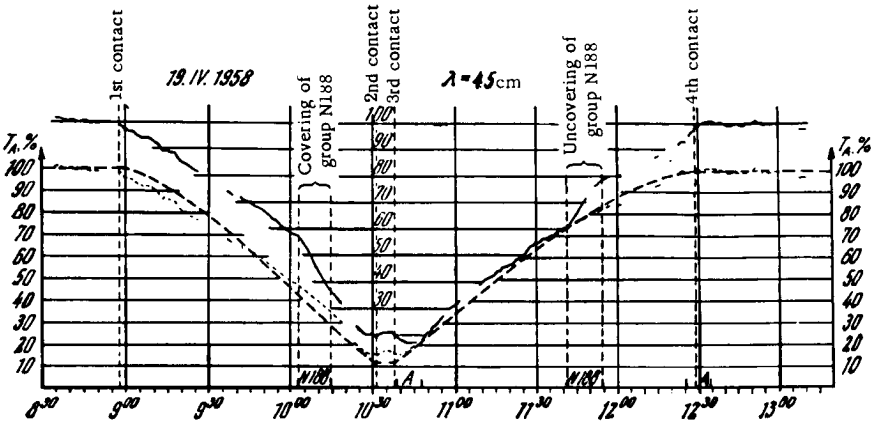


FIGURE 81

2. The sizes of local sources and the effective temperature T_{eff} of their radiations have been determined best from observations of solar eclipses. When the quiet sun is eclipsed by the moon, the radio flux subsides gradually and then gradually increases again after the phase of maximum eclipse. If a local source exists on the sun during the observation, then its occultation by the moon produces an additional change in the recorded flux, and this shows up clearly against the background of the monotonic variation due to the gradual eclipse of the undisturbed areas of the solar disk. An example of such a recording made during a solar eclipse is shown in Figure 81. Here the solid line shows the flux variation during the eclipse, the thick dashed line indicates the exposed area of the optically visible disk, and the fine dashed line shows the flux variation obtained after eliminating the effects of local sources. Since the velocity of relative motion of the moon is known, the size and flux of a local source can easily be estimated from the time and amplitude of the additional flux variation. Thus, we can also calculate its effective temperature

$$T_s = \frac{F}{\Omega_s} \frac{\lambda^2}{S_A 2k}, \qquad (32)$$

where S_A is the effective antenna temperature.

Eclipse observations were made repeatedly between the years 1942 and 1962. Some of the results obtained in these observations are listed in Table 11.

The values of T_b in the table are close to the electron temperature of the solar corona ($\sim 10^6$ deg), and the dimensions of the local radio sources are

close to the dimensions of optical spot groups. Figure 82 gives a comparison of these dimensions, as obtained during the eclipse observations on 19 April 1958 (a) and 15 February 1961 (b).

TABLE 11

λ, cm	Date of observation	Source dimensions in minutes of arc	T_b, °K
3.2	1.XI. 1951		$0.6 \cdot 10^6$
3.2	2.XII. 1956	0.5	10^6
3.2	11.IX. 1950		$2 \cdot 10^6$
3.2	19.IV. 1958	~3	$1.7 \cdot 10^6$
3.3	19.IV. 1958		$1.6 \cdot 10^6$
4.5	19.IV. 1958	~3	$1.9 \cdot 10^6$
5.1	19.IV. 1958		$3.6 \cdot 10^6$
10.7	23.XII. 1946		$1.5 \cdot 10^6$
10	1.XI. 1948		10^6
10	20.VI. 1955	4	10^6
4.5	15.II. 1961	1.2	10^6
9	15.II. 1961	0.5×1.14	$2.7 \cdot 10^6$
21	15.II. 1961	2×4.6	$2 \cdot 10^6$

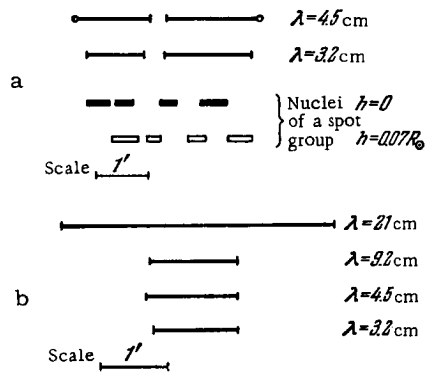


FIGURE 82

3. The heights of A-type radio sources are defined relative to the level of sunspot appearance in the photosphere. In order to determine the elevation of a radio source, we measure the velocity of motion of the source projection onto the solar disk during the rotation of the sun. Naturally, the greater the height h is, the larger will be the difference between this velocity and the velocity of migration of optical formations, since the radius of the sphere across which the radio source moves is greater than the radius of the photosphere (Figure 83). By measuring the difference in these velocities, we can find the height h; some of the results are summarized in Table 12. The facts listed above, that T_s is close to the coronal temperature, that F_s varies slowly, and that the source dimensions are close to the size of spot groups, indicate that the local radio sources may be coronal condensations located above large spot groups.

We have already observed that according to optical observations the electron densities in these condensations may be as much as 20 times the normal density at the same level in the solar atmosphere, while the electron

temperature may reach 10^7 deg. Examples of the arrangement of condensations above spot groups in the optical range are shown in Figure 84 (a is in visible light, and b is in the yellow line λ = 5694 Å) and in Figure 85 (the first and third frames from the top show the condensations as observed in visible light; the other two frames give the distributions of the lines λ = 5694 Å and λ = 5303 Å).

TABLE 12

λ, cm	$h/R_\odot$	h, km	Year of measurement	Antenna
2	0.06±0.07	42 000—50 000	1959	Knife-diagram radio telescope, Pulkovo, USSR (see Table 1)
3.2	0.07±0.02	50 000±14 000	1957—1958	
10				
7.5	0 05	35 000		Horizontal-base interferometer, Japan
21	0.035—0.1	20 000—100 000		Horizontal-base interferometer, Australia
60	0.05±0.02	35 000±15 000		
150	0.4±0.04	280 000	1949	Vertical-base interferometer, Crimea, USSR
200				

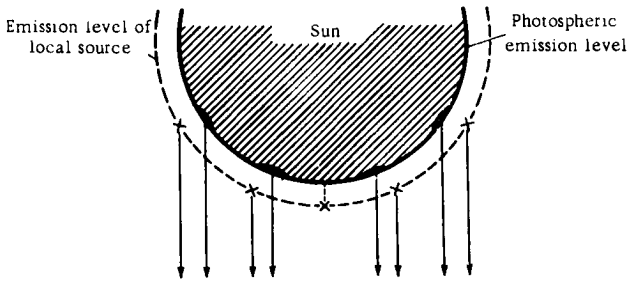


FIGURE 83

Optical observations show that the height of a coronal condensation above the photosphere is generally (0.05 to 0.1)$R_\odot$, which agrees well with the elevations found for radio sources. The radio sources, if they are located above spot groups, should thus rise before the sunspots and correspondingly should set later, as is actually observed. Radio-astronomical observations of the shifts in the center of gravity of the solar radio emission can thus be used to predict the emergence of spot groups. Such predictions are even easier to make using a large radio telescope whose high angular resolving power makes it possible to detect individual sources emerging from beyond the limb, regardless of the other sources already distributed over the face of the sun.

By applying the results of systematic observations (made in 1949) at λ = 10.7 cm, Waldmeier computed the radio emission of a condensation. He used the simplified model shown in Figure 86 (where $\varrho = r/r_\odot$) to obtain theoretical results which corresponded to observations of the randomly polarized radio emission of the local sources. According to his calculations, a single average condensation enhances the solar radio emission at a wavelength $\lambda \approx$ 10 cm by about 7 or 8%; at other wavelengths the effect is smaller.

Somewhat more detailed calculations, which utilized the results of eclipse observations, were carried out by Piddington and Minnett. They also took into account the influence of a 1500-gauss magnetic field. The results of these calculations also corresponded approximately to the observations, and moreover they explained the appearance of a polarized component in the emission of local radio sources. Thus, if we neglect for the time being the origin of the polarized component, the data at hand would appear to indicate that the emission of the coronal condensations forming above sunspots represents the principal constituent of the slowly varying component of the solar radio emission.

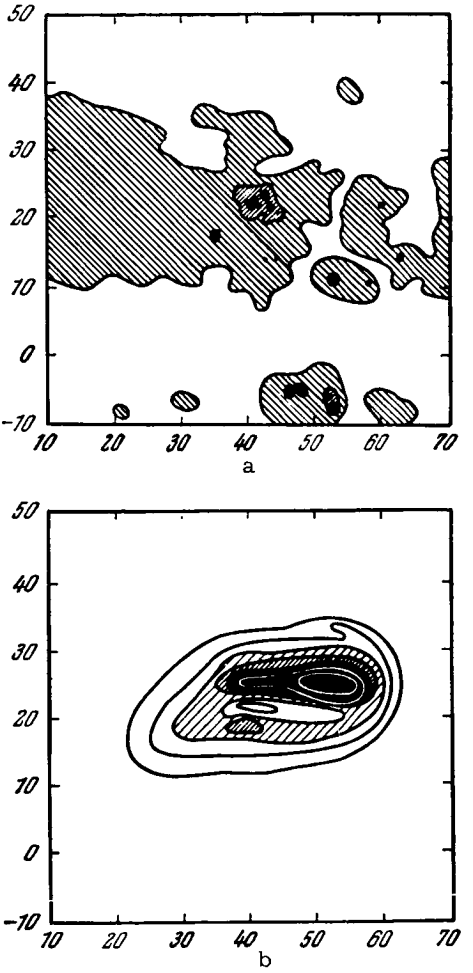


FIGURE 84

The polarization of radiation emitted by sources over spot groups. Let us now consider the last-mentioned property of local

radio sources, namely that they emit some circularly polarized component. This is a logical consequence of the fact that the local sources exist above sunspot groups, and thus are within the group magnetic field. We have already observed that magnetic fields are present in all sunspots, and that they can reach 4000 gauss in a large spot. A direct consequence of the location of an emission source in a magnetic field is anisotropy of the medium through which the radio waves propagate, and this in turn leads to the appearance of a circularly polarized component.

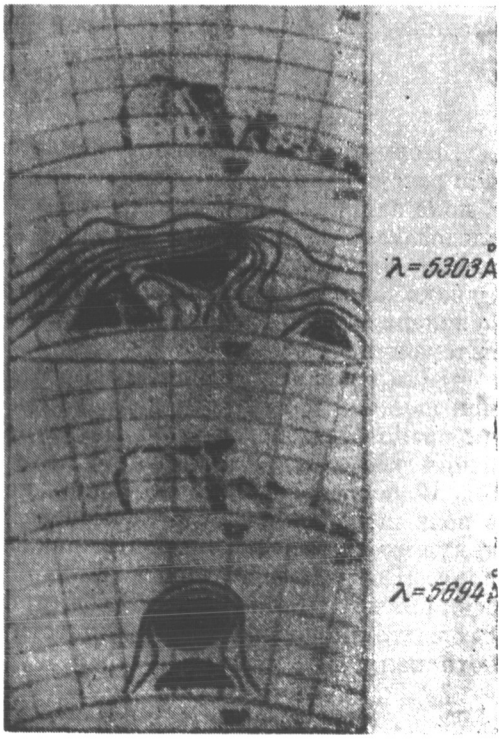


FIGURE 85

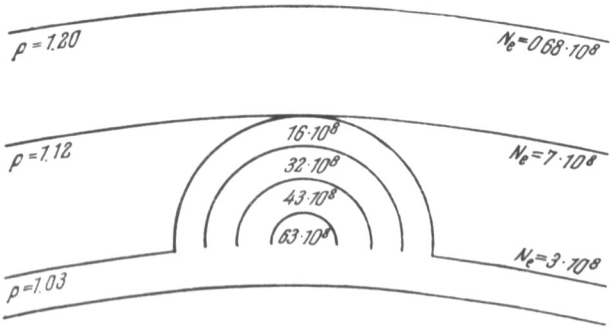


FIGURE 86

Let us now consider this phenomenon in more detail. First it will be useful to estimate the conductivity σ of the medium, since this is the property which determines the absorption of the propagating radiation. Actually, the electric power W consumed when a current flows through a conducting medium (ionized gas in this case) is given by the familiar formula

$$W=\frac{E^2}{2R}=\frac{E^2}{2}\sigma, \tag{33}$$

where E is the amplitude of the electric field of the radio wave, and R is the resistance of the medium.

The conductivity of the medium directly determines the absorption of the signal. On the other hand, this absorption can also be determined as the power lost by the radio wave in increasing the collision energy of the oscillating charges in the electric field, in comparison to the collision energy characteristic of a medium free from radio waves (when only thermal motion is present):

$$W^*=\frac{mv^2}{2}\nu N_e, \tag{34}$$

where m is the mass of a charged particle (in the solar atmosphere, this is the mass of an electron), ν is the number of collisions per second, N_e is the charge concentration per cm^3, and v is the velocity increment acquired by the charges in the electromagnetic field.

The quantity W^* calculated from (34) is defined per unit time (per second), so that it can be equated with the power W defined in (33). This makes it possible to relate the conductivity σ to the parameters characterizing the state of the solar atmosphere. Formulas (33) and (34) combine to give

$$\sigma=\frac{m}{E^2}v^2\nu N_e. \tag{35}$$

In order to find E and v, we merely substitute into (35) the following simple equations from mechanics and electrostatics:

$$mv=pt \quad \text{and} \quad p=Ee, \tag{36}$$

where p is the force acting on the charge e, and t is the time during which the force acts. Consequently, we find that the conductivity of the medium can be calculated from the formula

$$\sigma=N_e\frac{e^2}{m}\nu t^2. \tag{37}$$

From (35) it follows that the conductivity σ is determined by the mean velocity of charge displacement. Of course, this mean velocity will increase if the force p acts in one direction for a relatively long time. Therefore, for a sinusoidal variation of the field E at a frequency ω, the time t entering into (37) will be proportional to half the period. Consequently,

$$\sigma\sim N_e\frac{e^2}{m}\nu\frac{1}{\omega^2}\sim N_e\frac{e^2}{m}\nu T^2. \tag{38}$$

This relation is an approximate one, since we have neglected the influence which the collision of charged particles with other particles of the medium

has on the mean velocity of charge displacement or on the time during which the force acts in a single direction. This approximation obviously applies only if $\omega \gg \nu$. This approach, however, does not distort the description of the physical factors responsible for the polarized component, so that it will be adopted in the analysis that follows.*

In order to describe the propagation of radio waves in the solar atmosphere in the presence of a magnetic field, let us consider two particular cases: 1) the radio wave propagates parallel to the magnetic field, and 2) the radio wave travels along a perpendicular to the magnetic field.

In the first case, electric charges moving in the radio-wave field are acted upon by the Lorentz force, which is perpendicular to both the direction of motion and the direction of the magnetic field. A charge e moving through a field E will be deflected by this force, and the resultant motion (provided the velocity is constant) will be along a circular path. The frequency of angular revolution ω_H depends on the magnetic-field strength H:

$$\omega_H = H \frac{e}{m} . \tag{39}$$

According to (38), the period T during which the force acts upon the charge e in a single direction must be evaluated in order to estimate the conductivity σ. For this purpose, it is convenient to represent a linearly polarized oscillation as the sum of two component oscillations of frequency ω, which are circularly polarized in opposite senses.

Now it is clear that, if a component vector rotates in the same sense as the sense in which the charge revolves, then the period T during which the force p maintains a constant direction increases, and thus the conductivity σ becomes higher. In this case, in equation (38) we should substitute $\omega^* \quad \omega - \omega_H$ for ω, so that

$$\sigma_1 \sim N_e \frac{e^2}{m} \nu \frac{1}{(\omega - \omega_H)^2} . \tag{40}$$

In the particular case when $\omega = \omega_H$, radio waves are not transmitted through this section of the atmosphere, because of the strong atmospheric absorption (see (33)). The component which experiences increased absorption is often called an "extraordinary" wave.

The other component, whose vector E rotates in a sense opposite to the sense of charge revolution, will obviously have a higher frequency of force reversal, and the corresponding conductivity will be lower:

$$\sigma_2 \sim N_e \frac{e^2}{m} \nu \frac{1}{(\omega + \omega_H)^2} . \tag{41}$$

This component is often called the "ordinary" wave; it is absorbed less during its propagation. More detailed calculations give the following expression for the absorption coefficient χ in the presence of a longitudinal magnetic field:

$$\chi_{1,2} = \xi \frac{N_e^2}{T_e^{3/2}} \frac{1}{(f \pm f_H)^2} . \tag{42}$$

* In the wavelength range used for radio-astronomical observations of the sun, it is always true that $\omega \gg \nu$. It follows from (9) that this inequality is in very great danger of being violated only when T_e is small and N_e is large. However, we cannot assume here that $T_e < 6 \cdot 10^3$ deg and $N_e > 10^{10}$, although even in this case $\frac{\nu_{eff}}{\omega} < 10^{-2}$ for the longest waves ($\lambda \approx 50$ m).

This formula differs from (10) only in that f is replaced by $f+f_H$ for the ordinary wave (subscript 2) and by $f-f_H$ for the extraordinary wave (subscript 1).

It follows from the preceding that linearly polarized oscillations propagating through the ionized atmosphere of the sun parallel to the magnetic field are split into two components which are circularly polarized in opposite senses. These components are absorbed in different ways in the atmosphere, since the coefficient of absorption is a function of the sense of polarization. Therefore, when the radiation originates in a medium which is not very dense but has a substantial magnetic field, then one of the components (namely the extraordinary wave, whose χ is higher, as equation (42) shows, and thus whose optical depth is correspondingly greater) will predominate, and the entire radiation will have a certain net circular polarization, which increases with an increase in the magnetic-field strength H.

It is easy to show, using analogous arguments, the difference in the dielectric constants of the medium for the ordinary and extraordinary waves.

In the second case, when the radio waves propagate at right angles to the magnetic lines of force, oscillations whose vectors E are oriented along the magnetic field are not influenced by the magnetic field. Here, the charge motion will not be affected, since the Lorentz force is proportional to the sine of the angle between the directions of charge motion and of the magnetic field. However, for oscillations whose vectors E are perpendicular to the field H, charges moving through the radio-wave field at right angles to the magnetic field will interact with the latter; thus, the charge velocity and the absorption of these oscillations will increase somewhat. *

Consequently, if randomly polarized radio waves are transmitted through the solar atmosphere in the presence of a perpendicular magnetic field, then two groups of waves polarized in mutually perpendicular planes will have different intensities. An observer will thus receive radiation possessing partial linear polarization rather than completely random polarization.

The foregoing analysis was an approximate one ($\omega \gg \nu$), and it dealt with only two particular cases of relative orientation between the radio-wave propagation and the magnetic force lines. The properties of such propagation are discussed in more detail in specialized references such as /18/.

The properties of radio-wave propagation in a magnetic field are very significant with respect to the radiation from local radio sources located near sunspots, where the magnetic fields may be quite strong. It follows from the preceding that the radiation from these local sources should have a partial circular polarization; the sense of rotation of the polarization vector will depend on the direction of the magnetic field of the spot. It also appears that, as the spots migrate toward the solar limb, the degree of polarization will decrease (since the angle between the magnetic field and the vector changes), and partial linear polarization may appear. The first assumption is confirmed by observations.

Radio-astronomical observations first detected the circular polarization of local radio sources at longer wavelengths (λ = 310 cm), and then (as the equipment was improved) the polarization observations were extended into the region of shorter wavelengths. In the centimeter range, the polarized component may be as much as 45% of the total unpolarized flux of the source. At

* In the limiting case, when the rotation time of a charge in the magnetic field is equal to the period T (see above), the period $T^* \to \infty$ and $\sigma \to \infty$.

wavelengths shorter than 1 cm no polarization has yet been observed, possibly due to the insufficient sensitivity of the equipment.

The spectra of emission from sources over spot groups. One very important characteristic of any source is its spectrum. By applying relations (18) and (19), we can form an idea of the spectrum to be expected in the given case. If no magnetic field is present, then for shorter radio waves a condensation may have a small optical depth $\tau \ll 1$ and F_S will be approximately constant. As the wavelength increases, the optical depth of a condensation may become greater ($\tau \gg 1$), and then $F_S \sim \frac{1}{\lambda^2}$.

In the presence of a magnetic field, the absorption in a condensation increases with increasing wavelength. Actually, equation (42) can be written for simplicity in the following approximate form (for $f_H/f \ll 1$):

$$\chi_{1,2} \approx \xi \frac{N_e^2}{T_e^{3/2}} \frac{1}{f^2\left(1 \pm \frac{f_H}{f}\right)^2} \approx \xi \frac{N_e^2}{T_e^{3/2} f^2}\left[1 \mp 2\frac{f_H}{f} + 3\left(\frac{f_H}{f}\right)^2\right]. \qquad (43)$$

However, if the total flux is recorded, then as we have observed above, we must take into account that it is made up of two components circularly polarized in opposite senses (the ordinary and extraordinary waves):

$$F_\Sigma = \frac{F_1 + F_2}{2}. \qquad (44)$$

For an almost transparent source,

$$F_\Sigma = \frac{I^*}{2}\int (\chi_1 + \chi_2)\, dl = \frac{I^*}{2}\int \chi_\Sigma\, dl,$$

where

$$\chi_\Sigma = \chi_1 + \chi_2.$$

Then, by applying (44), we find that

$$\chi_\Sigma \approx \xi \frac{N_e^2}{T_e^{3/2} f^2}\left[1 + 3\left(\frac{f_H}{f}\right)^2\right]. \qquad (45)$$

In the presence of a magnetic field, the absorption in a condensation will be greater at longer wavelengths. Therefore, the spectrum to be expected for a thin source corresponds not to $F_S \approx$ constant, but rather to

$$F_S \approx \text{const}\left[1 + 3\left(\frac{f_H}{f}\right)^2\right]. \qquad (46)$$

Since for $\tau \gg 1$ we have $F_S \sim 1/\lambda^2$, it is possible to conclude that the spectrum $F(\lambda)$ has a maximum, which becomes more pronounced as the magnetic-field strength increases. Since all spot groups have magnetic fields, this dependence should apparently apply for all sources. It can be computed using an exact formula, provided we assume that the radiation emitted by a source is thermal in origin. Then, by applying (12) and (44), we find that

$$F_\Sigma = \frac{kT_e}{\lambda^2}\Omega_S\,(2 - e^{-\chi_1 l} - e^{-\chi_2 l}), \qquad (47)$$

where Ω_S is the solid angle subtended by the source, and l is the line-of-sight extent of the source. Calculations made according to (47) give the spectra shown in Figure 87 (here $a = kT_e\Omega_S$ and $E \approx 10^{-22}\frac{N_e^2}{T_e^{3/2}} l$).

The difficulties connected with making absolute radio measurements also hinder the construction of a spectrum from observational data. Therefore, the most reliable results are those obtained during solar eclipses, when the relative spectrum $\frac{F_S(\lambda)}{F_0(\lambda)}$ can be determined at various wavelengths. Also, by using the empirical expression for $F_0(\lambda)$ (see formulas (29) and (30)), we can convert from the results of observations to the absolute spectrum in a certain part of the short-wave range.

Such calculations were made for the observations of eclipses on 19 April 1958 and 15 February 1961; the resulting spectra are plotted in Figure 88.

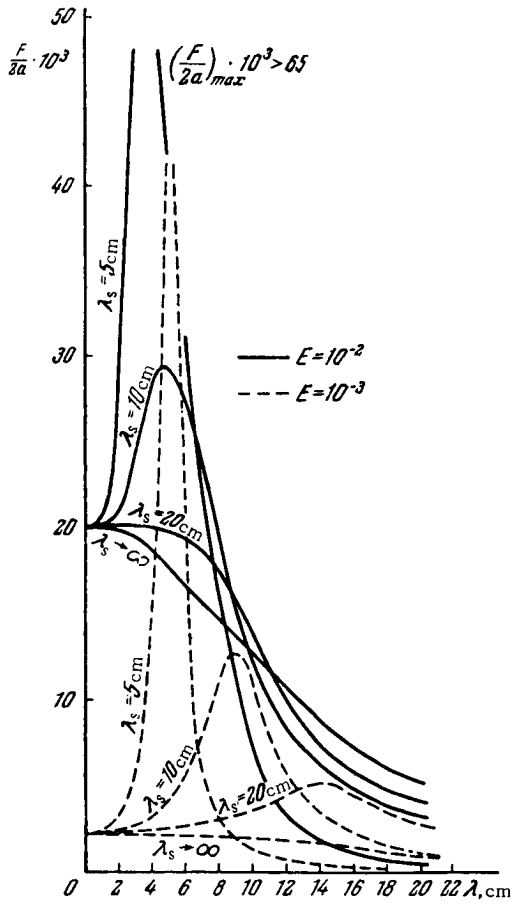


FIGURE 87

It follows from the preceding that the wavelength λ_0, at which the spectrum passes through a maximum, corresponds to a situation in which the optical depth of the condensation for the extraordinary wave increases to such an extent ($\chi_2 \to \infty$) that the source becomes opaque. If we compute the slope of the left part of the spectrum (where $\lambda < \lambda_0$) for $\lambda \ll \lambda_0$, or if we use the actual value of λ_0 (provided the electron concentration is not very large),

then we can find the magnetic field at source level from relations (45) or (47).

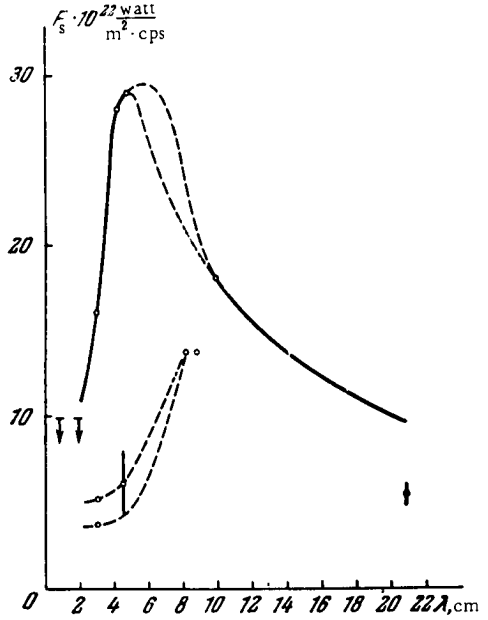


FIGURE 88

One consequence of the variation of the absorption coefficient χ with the magnetic field is a reduction of the observed radiation fluxes of the local sources as they migrate from the central meridian to the limb. This is due to the fact that the increment of χ determined by (44) applies to a longitudinal magnetic field only, that is, to cases when the source is located near the central meridian. For a transverse field (near the limb) the absorption is lower, so that both τ and F are proportionately smaller. All the preceding applies to the wavelength range $\lambda < \lambda_0$, where the optical depth of the sources is comparatively small. If, on the other hand, $\tau \gg 1$, then the radiation flux is determined by the product $T_e\Omega_s$. Here F_s may still vary with the source position, due to a variation in the apparent size of the source (which may take place for noncircular sources).

A comparison of the observed mean solar-radiation fluxes with the migration of spot groups across the solar disk shows that, on the average, the preceding dependences actually exist.* They are illustrated in Figure 89, where θ is the source distance from the center of the disk, and $i(\theta)$ is the decrement of the radiation flux; the dashed line corresponds to the case $i(\theta) = \cos\theta$.

The correlation between the optical and radio-astronomical characteristics of sources over spot groups. We can now return to the question of the correlation between the measured

* This fact should be taken into consideration during a statistical processing of the results of observations giving both the source flux F_s and the source position.

fluxes and the sunspot areas. A processing of the results of systematic observations has shown that the correlation coefficient between these quantities decreases with increasing wavelength. Once local radio sources have been identified with coronal condensations, however, this phenomenon is understandable, provided it is taken into account that decimeter waves are more characteristic of condensations than centimeter waves; the latter is true because condensations are not very transparent to decimeter waves but are almost transparent to centimeter waves. On the other hand, we know that condensations have longer lifetimes than visible spots do.

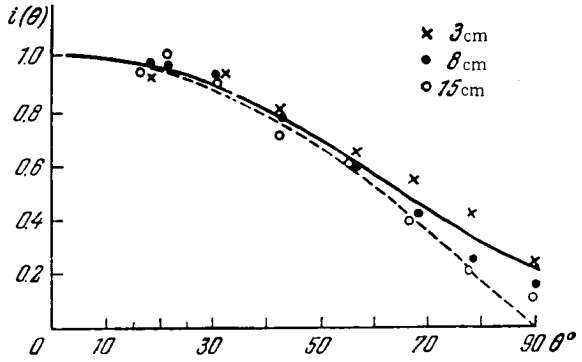


FIGURE 89

This optically established fact explains why local radio sources associated with sunspots may remain active even after the spots have disappeared. A statistical processing of the observation results indicates that some sources persist for almost two solar rotations after the disappearance of the sunspot. Therefore, it is clear that, when looking for a correlation between F_s and S_p, we must take into consideration not only the parameters of sunspots which are visible at the time of observation but also the parameters of sunspots which existed during previous rotations of the sun. Here, the highest correlation coefficient between optical and radio observations (for 25 cm $<$ $<\lambda<$ 50 cm) is obtained if we introduce a composite spot area $S_{com} = 0.4S_0 + 0.3S_{-1} + 0.25S_{-2} + 0.15S_{-3}$ (where the subscripts indicate the relative numbers of the solar rotations).

For λ = 10 cm, the composite spot area is best expressed in the form $S_{com} = 0.85S_0 + 0.55S_{-1} + 0.45S_{-2} + 0.35S_{-3}$. Here r = 0.92, as compared to the previously mentioned value of r = 0.78 (see p. 210) which was obtained using the total spot area at the time of observation.

In accordance with the preceding, the correlation coefficient for $\lambda < 10$ cm should decrease when we pass from "simple" to composite areas. Actually, for the displacement of the center of gravity of the solar radio emission measured at λ = 3.2 cm we have $r \approx 0.75$ in the first case and $r \approx 0.60$ in the second case (composite areas).

Due to the longevity of the local sources in the decimeter range, the undisturbed component can be determined from the minimum values of the solar radio flux only in those rare cases when the sun has remained "quiet" for a long time. Usually it is necessary to plot the graph $T_{av}(S_{com})$ and to find $T_{av\ min}$ as the intercept of the curve with the ordinate axis (Figure 90).

A statistical processing of quite long series of observed values of F_Σ, taking into consideration the various phenomena described above, may give T_{av} and the ratio $\frac{T_{s\,av}}{S_p} = \gamma$.* The results of such a processing carried out at NIRFI** are given in Table 13.

TABLE 13

λ, cm	T_{av} (1955–1956)	T_{av} (1956–1957)	γ	r for S_{com}	r for S_p
1.6	—	8 000	1	—	—
3.2	13 000	17 000	3	—	0.6—0.7
10	41 000	45 000	14	0.92±0.05	0.78±0.05
145	—	1 000 000	1000	0.60±0.06	0.28±0.03

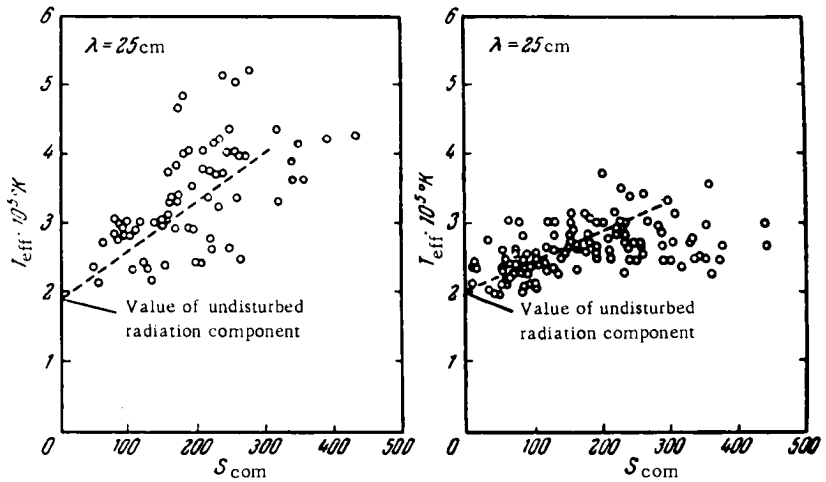


FIGURE 90

The almost total absence of the slowly varying component of disturbed radiation for $\lambda < 1$ cm is fully consistent with the proposed interpretation of the nature of this component in the centimeter range. This is true because the very-short-wave thermal radiation originates in the lower solar atmosphere, where the influence of coronal condensations is small; they are almost transparent in this range.

The observations with meter waves, however, contradict the foregoing results. In this range, the lifetimes of local sources do not exceed the spot lifetimes, and enhanced radiation is detected only when the sources are near the central meridian, that is, the radiation is highly directional (the directional-diagram width is generally close to 35° at λ = 60 cm, dropping to 15° at λ = 3.7 m). Moreover, the radio-emission intensity of the spot areas is so high (it is hundreds or thousands of times greater than the normal) that excitation by thermal means cannot be assumed in this range. If

* Here it is assumed that γ is constant. Recent investigations show that either γ is not constant or else $T_{s\,av} = \gamma S_p^n i(0)$ where $n > 1$.

** [Nauchno-issledovatel'skii radiofizicheskii institut (Scientific-Research Radiophysics Institute).]

we were to assume thermal excitation, then the effective source temperature would have to be about 10^{15} deg, which contradicts the results of optical observations.

The radio-astronomical data obtained so far thus indicate the possible existence of two different mechanisms causing the slowly varying disturbed component. In the centimeter and decimeter ranges, the thermal emission of coronal condensations is the main cause. In the meter range, the main contribution is apparently that of the radiation emitted during the motion of relativistic electrons in a magnetic field or else during plasma oscillations. This conclusion is backed up to a certain extent by the spectrum of the meter waves. Figures 91 and 92 give radio curves at various wavelengths for the first half of 1951, and Figure 93 shows the spectra for certain days (the broken lines in Figures 91 and 92 trace the undisturbed component). A distinct minimum is observed at $\lambda \approx 30$ cm, which can be regarded as the boundary of the spheres of influence of the two mechanisms referred to above.

The emission mechanism for sources over spot groups. The various observation results, together with their interpretation, led to the assumption that the main part of the slowly varying radiation component is thermal in origin. Many of the conclusions, however, were based on low-resolution observations, so that they naturally refer to some average source parameters. It is quite possible that the individual radiation sources differ greatly from one another with respect to their characteristics.

Some characteristics (the source area S, the brightness temperature T_b, etc.) were found more precisely from eclipse observations and with the aid of large radio telescopes. However, eclipse observations, which make it possible to distinguish between sources which are very close together, represent single-shot observations and cannot be repeated at will, while observations with large radio telescopes do not yet give reliable information on source spectra and on the polarization characteristics. Consequently, no conclusive description of the emission from local sources can be obtained by these methods.

Moreover, some recently discovered discrepancies in the observational results cannot be explained away just as due to a bremsstrahlung* mechanism /23/. To this class belong the following facts.

A. The magnetic-field intensities in local radio sources obtained from bremsstrahlung relations utilizing the results of polarization observations are considerably different from the values calculated according to simultaneously observed spectra. In the latter case, the field strength is 2 to 3 times higher.

B. The dependence of the radio flux on the source position relative to the center of the solar disk was found to be stronger than that predicted by the bremsstrahlung theory.

Furthermore, theoretical investigations of the emission of electrons slowed down in the magnetic fields of the sources /24/ have shown that this mechanism may introduce an appreciable contribution in the centimeter range. It is noteworthy that in this case radiation fluxes corresponding to the observed values originate in condensations whose electron concentrations

* Bremsstrahlung here refers to the radiation emitted when electrons are slowed down by repeated collisions. The radiation emitted when electrons are accelerated in a magnetic field changing their direction of motion is called magnetic bremsstrahlung. Both of these are equilibrium, that is, thermal, mechanisms.

are not greater than 10^8 to $10^9 cm^{-3}$, which is ten to one hundred times less than the concentrations required for bremsstrahlung.

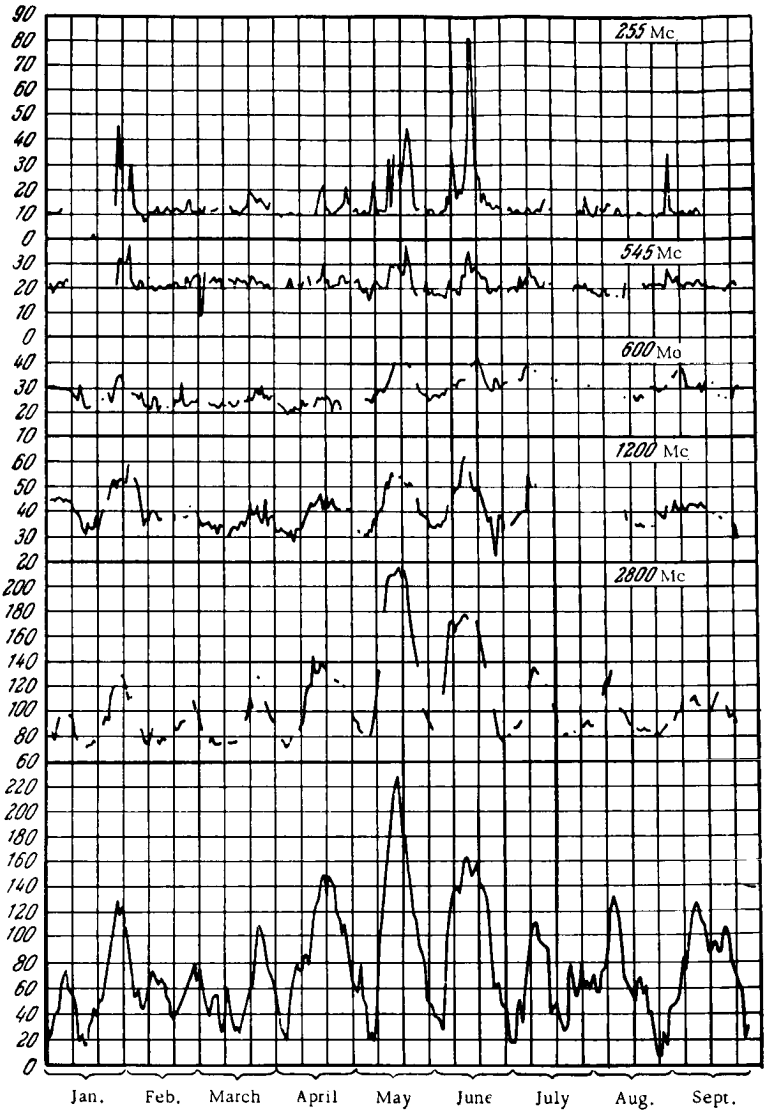


FIGURE 91

In view of these findings, the emission of *A* sources located in a magnetic field cannot now be explained only in terms of a bremsstrahlung mechanism. In a certain wavelength range, the main part of the emission is apparently due to the braking of electrons in regions of high magnetic-field intensity.

The theory of magnetic bremsstrahlung is still in the first stages of development, and no quantitative estimates can yet be made for comparison with the results of observations. The reader will therefore find further data on the problems of *A*-source emission only in periodicals.

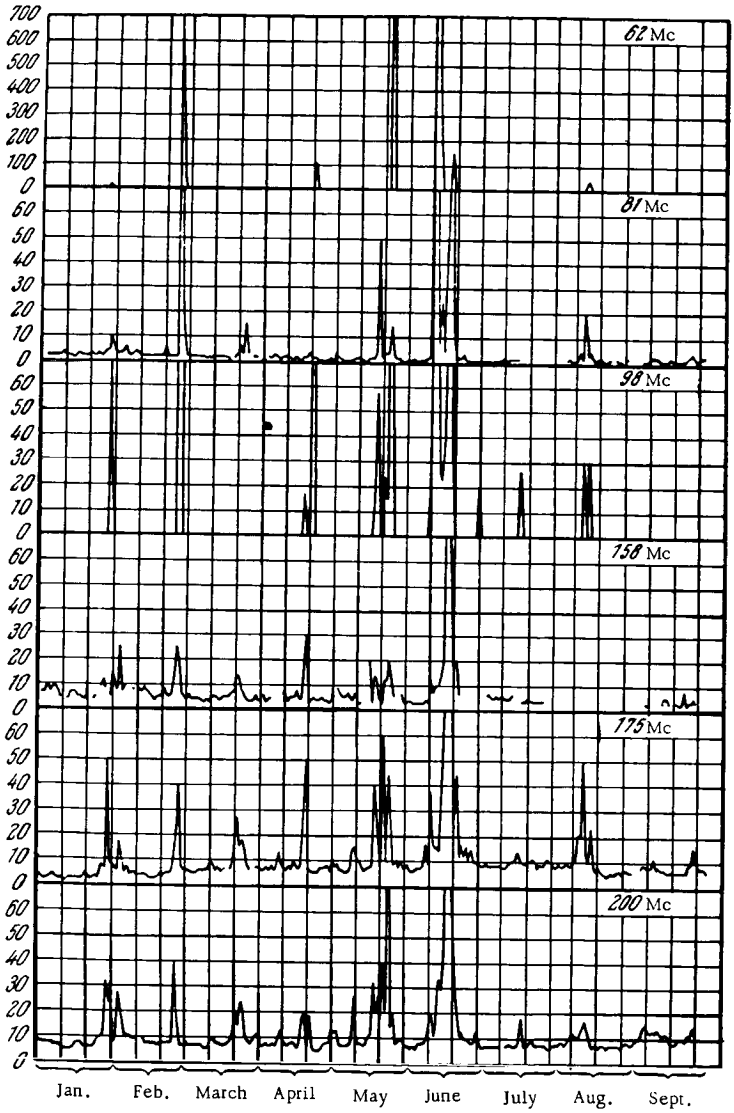


FIGURE 92

In addition to emission sources over sunspot groups, there are also sources over flocculi (B sources). Their brightness temperatures are substantially lower, but they are much larger in size, so that their radio fluxes can be recorded. During the eclipse of 15 February 1961, the spectrum of one such source was recorded (see Figure 94). It follows from

the nature of the spectrum ($F \approx$ constant) that the radiation in this case is generated by a bremsstrahlung mechanism, while the optical depth of the emitting layer is small. However, the scantiness of the observations precludes a more detailed description of such sources.

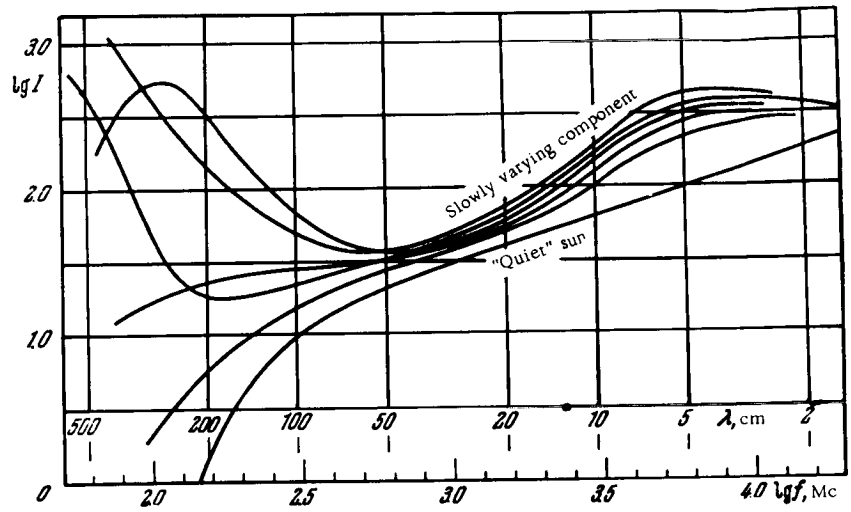


FIGURE 93

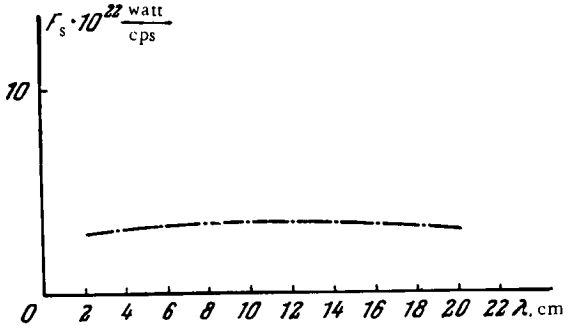


FIGURE 94

§59. **The rapidly varying disturbed component (radio bursts).** Disturbances of this kind are observed throughout the entire radio range, from millimeter waves to the longest meter waves. Some radio bursts are related to chromospheric flares, which are observable optically in visible light, in H_α emission, and at other wavelengths. However, in spite of the great importance of studies of these phenomena, the physical natures of both radio bursts and chromospheric flares are still far from being clarified. Nevertheless, they appear to be different facets of one and the same process, and their spectral differences are probably due to quantitative differences in the initial conditions.

Observations show that the radio bursts differ greatly in type. They may be either unpolarized or may have components possessing circular and linear polarization. Bursts may occur only in a limited wavelength range or their mean frequency may vary rapidly in time. They also vary in duration and intensity. Investigations made in years of low solar activity (1950-1951) have shown that the mean burst intensity increases with wavelength, often reaching values a thousand times greater than the undisturbed component in the meter range, then dropping to about 30% of the undisturbed component at $\lambda \approx 0.8$ cm, and rarely ever rising above the sensitivity threshold of the instruments at $\lambda \leqslant 0.4$ cm. The effective burst temperature in the meter range is often as high as 10^8 degrees, and in some cases it reaches 10^{12} degrees. A thermal emission mechanism thus seems to be unable to account for the physical nature of these bursts. The effective temperature of thermal radio emission naturally cannot be higher than the kinetic gas temperature, and these exceptionally high values are entirely inconsistent with the results of optical observations.

TABLE 14

Type of burst	Duration	Wavelength range	Polarization	Position on sun	Remarks
I (separate outburst, dense)	Seconds	Meter (each burst covers a narrow frequency band of some Mc)	Circular	Active regions	
II (large outbursts)	Minutes	Meter, decimeter (the band of a burst is some Mc wide)	Random, sometimes slightly circular	Regions of chromospheric flares	Band shifts at a slow rate, about 0.25 - 1 Mc/sec
III (isolated outbursts)	Seconds	Meter, decimeter (frequency band of burst is tens of of meters wide)	Random, sometimes slightly circular	Often identical with positions of chromospheric flares	Band shifts at rates up to 100 Mc/sec, observed more often than type II
IV	Minutes and hours	All ranges	Random and often circular	Identical with bright chromospheric flares at heights up to 10^5 km above photosphere	Sometimes source shifts rapidly at rates of 100 to 1000 km/sec. Often coincides with type-II burst. Rarely observed
IV-A	Minutes	Mainly centimeter and decimeter	Random and circular	Identical with bright chromospheric flares at heights up to 10^5 km above photosphere (height above photosphere around 50,000 km)	Rate of shift of source reaches 3000 km/sec. Frequently accompanies type-III bursts

The bursts are usually divided into different types according to the observed characteristics of their radiation. Inasmuch as no physical factors responsible

for their origin have been reliably established, this classification (shown in Table 14) is essentially phenomenological. The data in the table, which pertain to the burst spectra and to their variation in time, are illustrated in Figure 95.

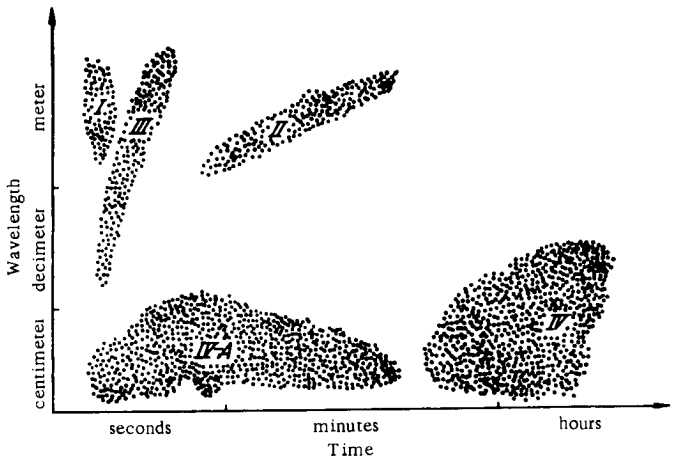


FIGURE 95

In addition to the types listed in the table, we sometimes encounter other types of rapidly varying solar radio components: radio storms (noise storms), classified in the preceding as a slowly varying component; gradual changes; U-bursts; inverted bursts; etc. Naturally, as observational material is further accumulated and interpreted, the characteristics in the table may vary.

The first three types of bursts (I, II, and III) differ from one another mainly in the rate of shift of the average frequency of the range in which they are observed. For the first type this shift is not appreciable, for the second it is about 1 Mc/sec, and for the third it can be up to 100 Mc/sec.

Observations using special equipment make it possible to achieve detailed studies of radio-burst spectra and their variations in time. In these observations, either the band of the received frequency is shifted continually or else a group of receivers are used, each tuned constantly to a different wavelength. In the most developed equipment of this type, the time constant for the measurements is less than 1/200 of a second.

The results of such observations, which have enabled the burst classification in Table 14 to be made, are illustrated by the dynamic spectra of bursts of types I, II, and III shown in Figures 96, 97, and 98 respectively. However, these spectra are difficult to interpret, without first making some assumption concerning the nature of radio bursts. Therefore, it is convenient to take as a working hypothesis the very plausible theory suggested in 1946 by Shklovskii. This theory explains the bursts of radio emission as an excitation of plasma oscillations in the solar atmosphere.

A quasineutral ionized gas, such as the gas constituting the solar atmosphere, is called in physics a "plasma." A plasma undergoes certain natural free oscillations, which arise during its excitation by external

sources. The frequency of these vibrations depends on the electron concentration in the gas and is defined by the following formula:

$$\omega_0 = \left(\frac{4\pi e^2 N_e}{m}\right)^{1/2}. \qquad (48)$$

A more detailed study indicates that (longitudinal) waves with both somewhat higher and somewhat lower frequencies than ω_0 can be propagated in plasmas.

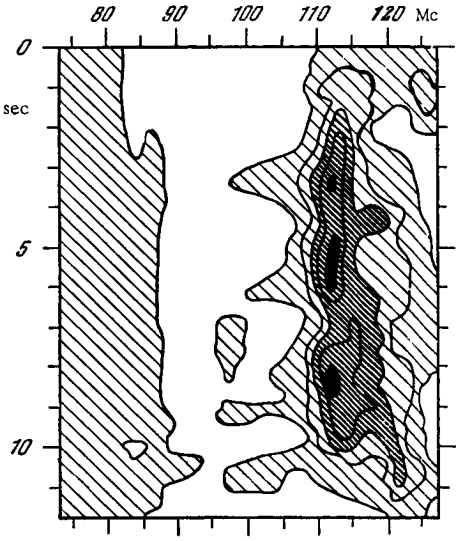


FIGURE 96

It is clear from a comparison of formula (48) with formula (7) that the value of ω_0 corresponds to the condition $n = 0$, that is, such oscillations can originate just at the boundary of a layer in which the propagation of vibrations of the given frequency is possible. Oscillations with $\omega > \omega_0$ will be propagated into the outer regions. Some difficulty arises in explaining theoretically the conversion of the energy of the longitudinal plasma oscillations into the radiation of transverse fields of electromagnetic waves. However, recently these problems have also begun to be solved /18/.

Using the hypothesis of "plasma oscillations," it is possible to interpret very graphically the dynamic spectra of radio bursts obtained during observations.

A type-I burst, which has no (real) frequency drift (Figure 96), can be explained as a plasma excitation in the condensed solar corona which usually exists near active regions. The duration of a type-I burst is generally only some tenths of a second. Calculations show that, for the conditions existing in a coronal condensation ($\nu \approx 5-20\,\text{sec}^{-1}$ and $T_e \approx 10^6-10^7$), the attenuation of normal plasma oscillations takes about this same amount of time, namely 0.1 to 6 seconds.

The exciting force for a type-I radio burst is as yet unknown. Since no frequency drift is observed, it may be assumed that this force is either localized in a small region or else is weak and gives rise to oscillations

only in the relatively dense condensations of the coronal plasma. At the same time, it may also be assumed that the physical conditions in some condensations are such that a reinforcement of the oscillations takes place, even one sufficient for the oscillations to be observed in the form of a radio burst.

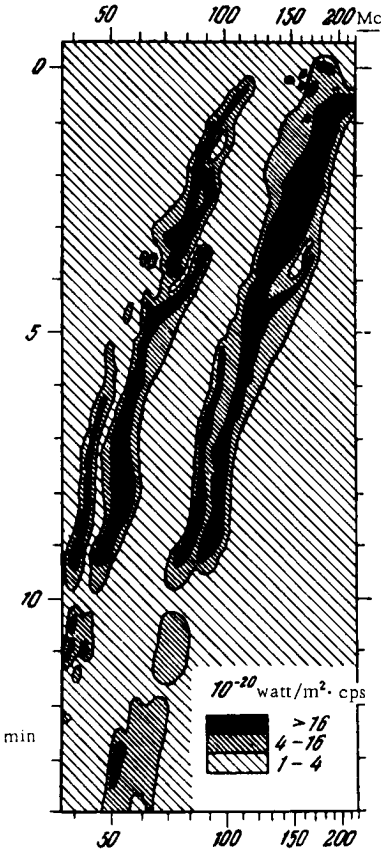


FIGURE 97

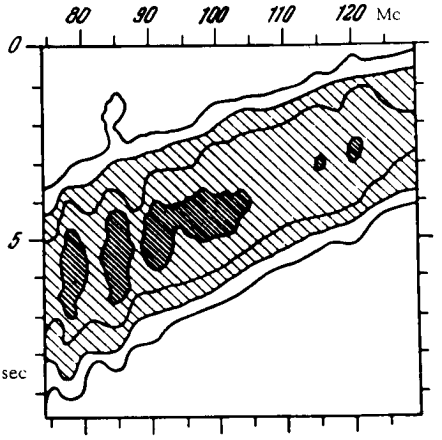


FIGURE 98

Type-I bursts correspond to $\lambda > 50 - 60$ cm. They constitute separate outbursts observable against the background of enhanced radiation from the active regions associated with spot groups located not very far from the center of the solar disk. Each type-I burst occupies a very narrow frequency band. Interferometer observations have established that all the type-I bursts associated with a single active region originate in the very same part of the sun, within an area less than three arc minutes in diameter. The degree of polarization is close to that of the slowly varying component of radiation from the same region, but it corresponds to the ordinary wave. These properties agree with the foregoing assumption that such bursts originate in relatively homogeneous and small-sized condensations of the coronal plasma. In particular, the presence and the nature of the polarization are explained in this case by the existence of strong magnetic fields in

the condensations, since this leads to an unequal absorption of radiation components with different senses of rotation. It follows from (40) that the extraordinary component of the oscillations is absorbed more strongly, so that in the radiation of the plasma oscillations the ordinary component will predominate.

Type-II radio bursts are encountered relatively rarely; even during periods of high solar activity, the average occurrence of these is once every 50 hours. These bursts are characterized by very high intensities, long durations, and large shifts in average frequency, which may be associated with the passage of an exciting agent through the interior of the solar atmosphere (Figure 97). During the motion of the latter outward from the photosphere, plasma oscillations originate at progressively higher levels, where N_e (and thus also ω_0) are lower. Actually, observations show that the locations of sources at different frequencies are different, and that they shift outward with an increase in wavelength. If we assume that the distribution $N_e(h)$ is known (for instance, if we take the distribution according to van de Hulst, see §57), then it is easy, using expression (48), to calculate the velocity of displacement of the exciting agent according to the rate of frequency drift. For type-II bursts this velocity is about 500 km/sec. Independently of this calculation, direct observations of the positions of bursts using interferometric methods also show that the regions in which type-II bursts originate shift with velocities of 500 to 3000 km/sec.

The value of the velocity suggests that the excitation of the plasma oscillations in this case is caused by matter ascending from the regions of chromospheric flares and eruptive prominences at velocities up to about 1100 km/sec. Similar velocities (1500 to 2000 km/sec) are possessed by the corpuscular streams causing sudden magnetic storms and polar auroras during the period from 22 to 30 hours after the appearance of a chromospheric flare near the central part of the solar disk. In relation to this, we should mention the good correlation between type-II bursts and geophysical phenomena on the earth, since this also supports the previously mentioned consideration concerning the relation between corpuscular streams and radio bursts of this type.

Finally, a very important property of type-II bursts (and also of type-III bursts) is that a burst at a frequency of $2\omega_0$ is often (in more than 60% of the cases) observed simultaneously with a burst at a frequency ω_0. As shown in Figure 97, the development of the phenomenon at $2\omega_0$ even follows in detail the process at ω_0. The similarity of these spectra and the frequency ratio compel us to assume that the two bursts originate in a single medium, possessing moreover a considerable nonlinearity of its parameters. *

In relation to this, one fact is very interesting, namely that the frequency ratio is close to, but always slightly less than, two (it is about 1.95). The explanation of this may be that oscillations in a band occupied by a burst with average frequency ω_0 (especially when $\omega \approx \omega_0$) are absorbed much more than those in a band with average frequency $2\omega_0$. Here, as noted previously, the more low-frequency part of the band around ω_0 is absorbed more strongly than the higher-frequency part (Figure 99). For this reason, the average frequency ω_0^* recorded during observation shifts somewhat toward higher frequencies, and so the ratio $2\omega_0/\omega_0^*$ becomes a bit less than

* In a medium whose parameters depend on the intensity of the processes taking place in it (nonlinear medium), harmonic components with multiple frequencies will appear during the propagation of the oscillations.

two. The high absorption in regions where radio bursts appear, in particular, may account for their lower intensity at shorter wavelengths. In this range, the levels where radio-wave absorption is great ($\tau \approx 1$) are located considerably above the levels $n = 0$, in which the plasma oscillations originate. Consequently, the radio bursts originating in this layer will be absorbed strongly as they pass outward.

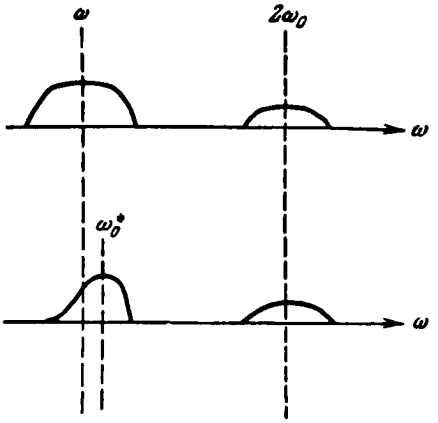


FIGURE 99

It is also very interesting that the plasma-oscillation theory predicts a splitting of the average frequency, corresponding to Zeeman splitting. A similar situation is sometimes observed during recordings of the dynamic spectra (see, for example, Figure 97, in the interval of time from 7 to 9 minutes). If in the future this relation is confirmed, then dynamic-spectra studies will make it possible to specify the magnetic fields in coronal regions located high above active regions of the sun.

Type-III bursts have the same general nature as type-II bursts, and all the foregoing applies to them as well (see Figure 98). The difference between the two types is just that for type-III bursts the rate of change of the average frequency is very high. Thus, if we assume that they are also caused by coronal oscillations of the plasma type, then the calculations described above result in a linear velocity of the exciting agent of about 30,000 to 100,000 km/sec (0.1 to 0.3 times the velocity of light). Interferometric measurements have verified this high velocity. Streams of particles at these velocities make up the soft component of cosmic rays of solar origin. Therefore, attempts are now being made to find a relation between these two phenomena, but no definite results have as yet been obtained.

In type-III bursts a slowing down of the frequency variation during the burst is often observed. This is probably a consequence of the negligible value of the electron concentration in the outer layers of the corona. Since the plasma-oscillation frequency ω_0 is determined by the overall quantity of electrons, both in the plasma and in the exciting flux (see formula (48)), therefore if the latter is constant the total concentration in the rarefied outer corona, and thus also the frequency, will remain almost invariable.

The duration of a type-III burst is from 3 to 15 seconds, and its intensity is hundreds of times greater than the level for the undisturbed sun. Just as bursts of spectral type II are associated with chromospheric flares, type-III bursts may have a second harmonic, which indicates the existence of nonlinear phenomena in the regions where the radiation originated.

A comparison of type-III radio bursts with optical phenomena on the sun shows that more than 60% of the bursts are associated with chromospheric flares having visible areas of more than 20 millionths of the area of the solar disk.

Type-IV bursts are observed relatively rarely, and they may have very different durations, ranging from a few minutes to several hours; in addition, they cover a wide spectrum of frequencies. The brightness "temperature" of the emitting regions for these bursts may be as much as 10^{12}. This type appears most often after strong photospheric flares and type-II bursts.

Contrary to bursts of types II and III, type-IV bursts do not have positions on the solar disk which depend on the frequency. The existence of such a dependence for bursts of types II and III agreed well with the explanation of these bursts as a result of plasma oscillations. Consequently, for type-IV bursts such an explanation will be less plausible.

Measurements of the displacement velocities of the regions in which type-IV bursts originate indicate values up to 5000 km/sec. The dimensions of these regions are considerably greater than those for the bursts discussed above.

Type-IV-A bursts differ from the other types mainly in that they appear in the centimeter wavelength range, usually precede other bursts (if other bursts appear at all), are observed over a wide band of frequencies without much frequency drift, and finally are very closely related to chromospheric flares. The systematic study of these bursts has begun only quite recently, but already the following interesting results have been obtained.

A. Systematic studies of the slopes of the spectra of type-IV-A bursts show that in most cases when the burst goes through its maximum a thermal radiation mechanism is implausible and that the radiation in this case is better explained by a magnetic-bremsstrahlung mechanism. Certain properties of the observed values of the spectral slope near $\lambda = 3.04$ cm imply the presence of monochromatic emission of excited hydrogen in the burst vicinity. Other observations of spectra over a wide range of centimeter and decimeter wavelengths make it possible to assume that for some bursts, especially after the maximum, a thermal radiation mechanism predominates.

B. Systematic observations of the burst-flux spectra over a still wider wavelength range, including meter waves as well, indicate that these spectra have, as a rule, two maxima. One maximum is around 6 or 7 cm, and the other is in the meter-wave region. The minimum separating these corresponds to a wavelength of approximately 0.5 to 1 meter. This property, which is reminiscent of a spectral property of the slowly varying component, agrees with the previously given classification of emission mechanisms: for bursts of types I, II, and III, observed in the meter range, the determining factors are the oscillations of the excited plasma; for bursts of types IV and IV-A, observed at shorter wavelengths, a magnetic-bremsstrahlung emission mechanism is more likely.

C. A close correlation has been observed between strong, prolonged bursts of types IV and IV-A, on one hand, and the subsequent incidence at the earth of powerful streams of protons with energies of 10 to 50 Mev. The production of the latter in the solar atmosphere is associated with bright chromospheric flares. The proton streams reach the earth after a delay of anywhere from one to several dozen hours. Curvature of the proton path from sun to earth exists due to the presence of magnetic fields along the way. Consequently, the highest probability for proton incidence at the earth will exist for streams originating in the western part of the solar disk.

The above facts lead us to hope that observations of type-IV-A bursts will make it possible to predict the appearance of ionizing radiation near the earth. This will be very valuable, for example, in order to guarantee the safety of manned space flights. The foregoing results are still only tentative, and they require a careful verification; however, they indicate what interesting problems and possibilities are involved here.

The hypotheses concerning the emission mechanisms have not yet been finally verified by observational results, in spite of the plausibility of explaining some of the phenomena in terms of plasma oscillations and others in terms of the braking of relativistic electrons in magnetic fields. One reason for this, namely the incompleteness of the theoretical development of these problems, prevents us from making reliable calculations which agree with the observations and also from predicting phenomena which were previously not discovered during the observations.

The radio development of such studies, however, makes it possible to anticipate a rapid and highly stimulating progress along these lines in the immediate future.

REFERENCES FOR CHAPTER IX

1. GINZBURG, V. L. — Uspekhi fizicheskikh nauk, 32: 26. 1947; Uspekhi fizicheskikh nauk, 34: 13. 1948.
2. GETMANTSEV, G. G. — Uspekhi fizicheskikh nauk, 44: 527. 1951.
3. RYLE, M. — Uspekhi fizicheskikh nauk, 46: 508. 1952.
4. Problemy sovremennoi fiziki (Problems in Modern Physics). — Ser. 3, No. 15, 1951; No. II, 1953.
5. SHKLOVSKII, I. S. Radioastronomiya (Radio Astronomy) — Gostekhizdat. 1955.
6. LOVELL and CLEGG. Radioastronomiya (Radio Astronomy) — Izdatel'stvo inostrannoi literatury. 1953 [Russian Translation from English].
7. KUIPER, G. P., Editor. Solnechnaya sistema (The Solar System), Vol. I, Solntse (The Sun). — Izdatel'stvo inostrannoi literatury, 1957 [Russian Translation from English].
8. PAWSEY, J. L. and R. N. BRACEWELL. Radioastronomiya (Radio Astronomy). — Izdatel'stvo inostrannoi literatury. 1958 [Russian Translation from English].
9. SHKLOVSKII, I. S. Fizika solnechnoe corony (Physics of the Solar Corona). — Fizmatgiz. 1962.
10. WALDMEIER, M. Ergebnisse und Probleme der Sonnenforschung. — Leipzig. 1955.
11. ZHELEZNYAKOV, V. V. — Uspekhi fizicheskikh nauk, 64: No. 1. 1958.
12. GURZADYAN, G. A. Radioastrofizika (Radio Astrophysics). — Izdatel'stvo Akademii nauk, Arm SSSR. 1956.
13. KHAIKIN, S. E. — Radiotekhnika. 10: No. 47. 1955.
14. Trudy V soveshchaniya po voprosam kosmogonii (Proceedings of the Fifth Conference on Problems of Cosmogony). — Izdatel'stvo Akademii nauk SSSR. 1956.
15. Radioastronomiya, Parizhskii simpozium 1958 (Radio Astronomy, the 1958 Paris Symposium). — Izdatel'stvo inostrannoi literatury. 1961.
16. Izvestiya Glavnoi astronomicheskoi observatorii, No. 164 (radioastronomiya). 1960.
17. Izvestiya Glavnoi astronomicheskoi observatorii, No. 172 (radioastronomiya). 1963.

18. GINZBURG, V.L. Rasprostranenie elektromagnitnykh voln v plazme (The Propagation of Electromagnetic Waves in a Plasma). — Fizmatgiz. 1960.
19. PIDDINGTON, J.H. —Proc. Roy. Soc. 203A: 417. 1950.
20. SHKLOVSKII, I.S. and E.V. KONONOVICH. — Astronomicheskii zhurnal, 35: 37. 1958.
21. TROITSKII, V.S. and N.M. TSEITLIN. — Izvestiya vuzov, Radiofizika. 4: 3, 393. 1961.
22. PANOVKIN, B.N. — Astronomicheskii zhurnal, 34: 505. 1957.
23. MOLCHANOV, A.P. — Astronomicheskii zhurnal, 38: No. 5, 849. 1961.
24. ZHELEZNYAKOV, V.V. — Astronomicheskii zhurnal, 39: Issue 1, 5. 1962.

PART II

THE PLANETARY SYSTEM

N 67-15494

A. V. MARKOV

Chapter X

THE MOON

§60. **The moon as a celestial body.** A correct knowledge of the nature of the lunar surface is particularly important today due to the prospective landing of men on the moon, an event which will occur in the very near future.

The diameter of the earth's satellite, the moon, is 3476 km, and its mass is 1/81.53 of the terrestrial mass. The mean density of the moon is therefore 0.66 of that of the earth, or 3.33 times the density of water. Since its mass is small, the moon cannot retain a noticeable atmosphere. A study of the figure of the moon shows that this satellite is a nonequiaxial ellipsoid with its major axis pointing earthward; this is due to the influence of the tidal forces set up by the pull of the earth. The difference between the axial lengths, however, is slight, amounting to only a few hundred meters.

As the moon rotates about its axis, which at present takes 27 days, 7 hours, 43 minutes, and 11 seconds, a time also equal to the sidereal period of revolution of the moon about the earth, the longest axis of the lunar ellipsoid is at all times directed earthward. Therefore, one side of the moon always faces the earth. However, due to the linear librations in longitude and latitude, 0.59 of the lunar surface is visible from the earth.

The almost absolute immutability of the spots on the lunar disk led ancient observers to conclude that the moon moves through space in such a way that one of its hemispheres always faces the earth. Thus it was obvious that the rotation period of the moon about its axis must equal the period of lunar revolution about the earth. With the advent of telescopic observations, the phenomena known as librations were discovered; these are small deflections of the lunar sphere about axes perpendicular to the line joining its center with the earth.

In 1610 the libration in latitude was discovered. This "rocking" motion is caused by the inclination of the equatorial and orbital planes of the moon with respect to the ecliptic. The libration in latitude is manifested by the fact that the north and south lunar poles become alternately visible to an observer on the earth. The selenographic latitude of the apparent center of the moon varies through an angle of ±6°.8 as a result of this.

In 1654 the lunar libration in longitude was discovered; here, the selenographic longitude of the apparent lunar center varies through ±7°.9. This libration is caused by the fact that, whereas the rotation of the moon about its axis is uniform, its motion in the elliptic orbit about the earth is, in accordance with Kepler's second law, nonuniform.

In 1693 D. Cassini formulated the following three laws governing the axial rotation of the moon.

I. The moon undergoes direct rotation with constant angular velocity about an axis maintaining a fixed position in the lunar sphere. The period of axial rotation is exactly equal to the sidereal period of revolution of the moon about the earth.

II. The equatorial plane of the moon makes a constant angle of 1°32' with the plane of the ecliptic.

III. The axis of the ecliptic, the axis of the lunar orbit, and the rotation axis of the moon are always coplanar, the axis of the ecliptic being located between the two other axes. As a result, it follows that the ascending node of the lunar equator always coincides with the descending node of the lunar orbit. The nodes of the lunar orbit continuously move along the ecliptic, completing one revolution in 18.6 years. From Cassini's laws it follows that the nodes of the lunar equator must move along the orbit with this same period, and consequently the rotation axis of the moon will describe a closed circuit in space during that time. This motion of the lunar axis is not unlike the precession of the earth's axis.

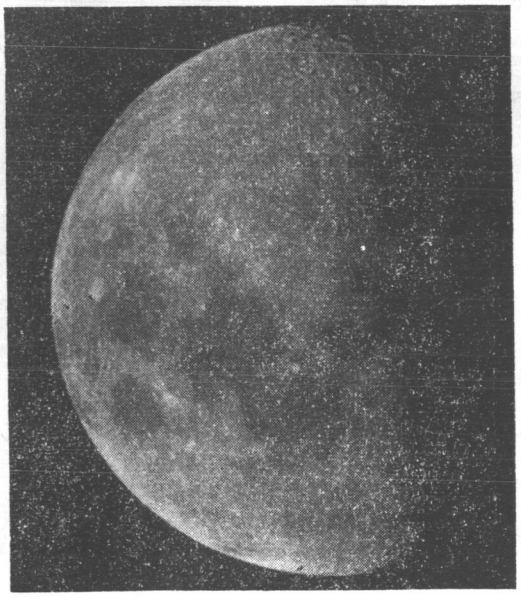


FIGURE 100

Cassini's laws describe the so-called optical libration of the moon. However, subsequent studies of the moon have established the existence of another, smaller, nonuniformity in the rotation of the moon about its axis. The latter, which is known as the physical libration, is a real oscillation of the axis which does not exceed 0°.02. The physical libration is divided into free and forced components. The free component shows up as a rocking motion of the moon about a mean position, and it may be due to a certain difference between the rotation periods of the moon at present and in the past. The forced libration is due to a periodic deviation of the

largest lunar axis from the earth-moon line, as a result of the optical libration.

In its structure, the surface of the moon is quite different from the surfaces of all the planets in the solar system except Mercury. The lunar surface is highly irregular. The more extensive, relatively light, part of the moon has a dense cover of mountain chains (up to 8 km high), ring-shaped mountains and craters (up to 230 km in diameter), and also fissures [rilles]. The rest of the lunar surface is covered with somewhat darker, sunken plains. Most of these plains are roughly circular, and they sometimes reach 800 km in diameter. These plains, erroneously called maria (or seas) by the first observers of the moon, are often encircled by high mountains. As Figure 100 (a photograph taken on the tenth day after new moon) shows, the lunar topography shows up particularly well near the terminator, where, due to the low height of the sun above the horizon, the surface irregularities are sharply accentuated by their shadows.

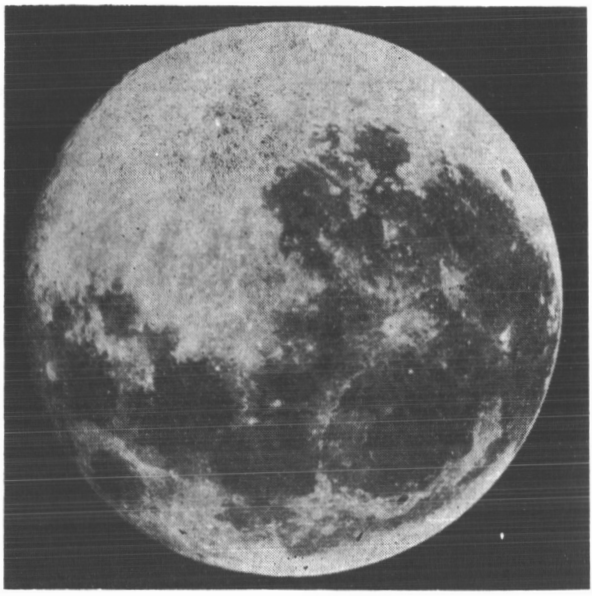


FIGURE 101

Near full moon no visible shadows are cast by the mountains, and the apparent brightness contrast of the lunar surface is then determined by the reflectivity differences between various areas. At this time a particularly prominent feature on the moon is a series of systems of bright rays, diverging almost radially from a few dozen of the ring-shaped mountains, whose walls then appear to be very bright (Figure 101). Some of the bright rays can be traced over distances as great as 1/4 of the lunar circumference. It should be added that these rays, without deviating from their general direction, pass through both "maria" and mountainous regions of the moon. However, a detailed study of the rays shows that they generally

consist of a multitude of individual streaks, which only on the whole follow the general direction of the ray. In some cases individual large rays are found to extend along a tangent to the outer wall of a crater, rather than from its radius. The rays have not been observed to influence the topography of the moon.

§61. Photometric and polarization properties of the lunar surface. The unique structure of the moon's topography influences the reflection of light from the lunar surface. Measurements of the integrated stellar magnitude of the moon at various lunar phases have shown that the mean spherical visual albedo of the moon (see Vol. I, Chap. XXII) is 0.07 and also that the lunar brightness drops with the phase angle α much more rapidly than is indicated by Lambert's law for a smooth, diffusely scattering sphere. Brightness measurements of individual sections of lunar "maria" and mountains obtained for different phase angles /1, 2, 3/ show that without exception all the moon's features reach maximum brightness (for an earth-bound observer) not when the sun is highest above their respective horizons, but rather at full moon, when no shadows are observed on the lunar surface. This indicates that all parts of the moon's surface are covered by small irregularities which cannot be distinguished through a telescope.

The reflectivity $\varrho = \frac{\pi B}{E}$ (where B is the brightness and E is the illumination) at full moon ($\alpha \to 0°$) varies in the "maria" from 0.05 to 0.08, in the craters from 0.06 to 0.16, and in the lunar rays from 0.10 to 0.21. Measurements of ϱ have shown that the largest area on the surface of the moon which is continuously covered with rocks and which has $\varrho > 0.21$ is 368 km^2 (100 square seconds of arc) in area.

Studies of color contrasts on the lunar surface, using light filters, colorimeters, and spectroscopic equipment, have shown that on the average the moon is reddish-brown in color. The integrated color index of the moon, using the international system, is $+0^m.35$ greater than the color index of the sun.

Some data on the reflectivity of lunar details and on the difference Δ between their color indexes and the solar color index are listed in Table 1. This table also gives some comparative data for terrestrial rocks and for meteorites. It is found that basic rocks are closer in reflectivity and color index to the lunar surface than acidic and sedimentary rocks are. Meteorites are closer to the moon in reflectivity but they have a more bluish color in some cases /4/.

Photometric studies of lunar details by means of spectroscopic equipment have been made in Germany, the Soviet Union, and America. Whereas the visual spectrophotometric samples of Wilsing (Figure 102) for wavelengths from 450 to 650 mμ confirmed that sedimentary rocks and granite are redder than lunar formations, still they also showed that basalt rocks are closer in spectral reflectivity to the lunar "continents" and that lavas are closer to lunar "maria."

A comparison of Wilsing's data with data on the average spectral reflectivity of the entire moon /5/, recently obtained (Figure 103) by photoelectric means for $\lambda\lambda$ 320 -550 mμ, indicates the following regularities. First, the slight drop in reflectivity and the subsequent increase over the range $\lambda\lambda$ 460 -530 mμ (particularly for the "continents") turns out to be an actual feature of the lunar spectrum. Second, a slight drop in the average reflectivity of the entire moon between 420 and 360 mμ (in addition to the general

downward trend from λ = 480 mμ to λ = 320 mμ), which shows up using the new technique, is also definitely real. This leads to the conclusion that the lunar surface may be covered with silicates (SiO_2).

TABLE 1

Lunar formations	ρ %	Δ	
"Maria"	5—8	from +0^m.40	to +0^m.36
"Pali"	9—10	+0.23	+0.28
"Continents"	8—12	+0.12	+0.46
Craters	6—16	+0.20	+0.48
Bright rays	10—21	+0.10	+0.48
Terrestrial rocks			
Limestone	5—80	from 0^m.13	to +1^m.52
Sandstone	6—54	+0.03	+1.52
Clays, shales	12—50	-0.24	+1.52
Granite, acidic rocks	4—70	-0.25	+0.99
Basalt, basic rocks	6—28	-0.31	+0.95
Meteorites	4—48	from -0.16	to +0.36
Fusion crust	2—17	-0.10	+0.38

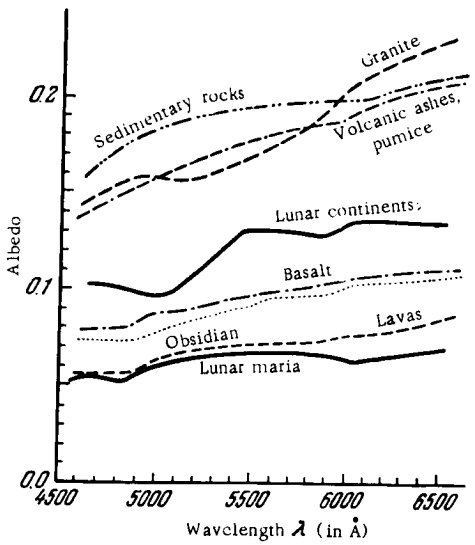


FIGURE 102

Spectrophotometric studies of a large number of individual lunar details were undertaken from 1957 to 1960 at three Soviet observatories. At Pulkovo, T. A. Polozhentseva (/6/, p. 148) studied 10 "continental" and "marine" areas in comparison with Sinus Iridum. In addition, N. P. Barabashov, A. G. Chekirda, V. I. Ezerskii, and V. A. Fedorets (/6/, pp. 148-150)

carried out a similar study using the spectrum of Mare Imbrium as a reference. Finally, V.G. Teifel' /7/, at Alma Ata, compared the spectral reflectivities of 261 small lunar objects in 11 regions with the reflectivity of Mare Vaporum.

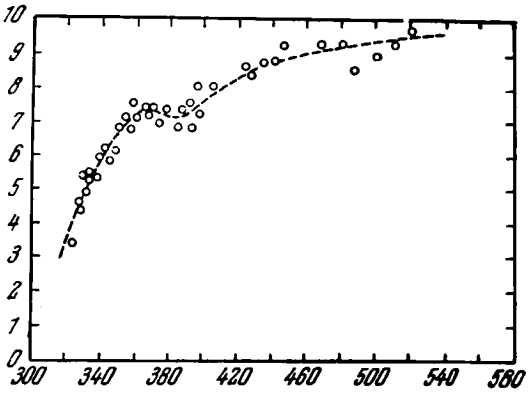


FIGURE 103

After Polozhentseva processed her small Pulkovo series, which had been obtained back in 1957, she came to the following conclusions concerning the color and spectral characteristics of the lunar areas studied:

1) the color contrasts between the lunar features considered are slight; they range between 10 and 15%;

2) the "continental" areas considered are as a rule redder than the "marine" areas.

The more extensive Kharkov series led Barabashov and his co-workers to the same conclusions. Finally, Teifel', on the basis of the processing of his large series, also confirmed the presence of small color contrasts on the moon. He furthermore compared the spectrum of his reference area (Mare Vaporum) with the spectrum of α Aurigae (G0) by the method of absolute spectrophotometry; within the margin of error introduced into these measurements by the variation of the spectral transparency of the atmosphere, he obtained the color indexes C of the lunar details included in his catalog.

Assuming for α Aurigae a color index of $+0^m.82$, Teifel' established that the color indexes of all the lunar details measured by him vary over a narrow range from $+0^m.76$ to $+0^m.97$, the reference area (Mare Vaporum) having an index of $C = 0^m.77 \pm 0^m.05$. The color of a lunar detail appears, according to Teifel' (and also according to other authors), somewhat redder as the albedo increases. Small lunar details did not show essentially greater color contrasts than extended formations. A comparison of the average indexes obtained by Teifel' for the darkest ($0^m.76$) and the lightest ($+0^m.97$) areas, and also of their spectral characteristics /8/, with previous data obtained using different methods is given in Table 2.

The polarization of the light reflected by the entire moon and by individual lunar formations has been studied in America, France, and the Soviet Union

/6/. These studies show that the degree of polarization $P = \frac{I_1 - I}{I_1 + I}$ of moonlight, where $I_1 - I$ is the intensity of the polarized component and I is the intensity of the unpolarized component, varies with the phase angle.

TABLE 2

Author	Year	Method	Color index	Spectral class
Russell	1916	Difference of photographic and visual magnitudes of moon	$+1^m.18$	—
Pettit	1925	Red-blue ratio	+0.92	
Barabashov	1925	Photographic photometry	—	F9—K0
Rougier	1936	Photoelectric method	+1.11	—
Radlova	1938	Visual colorimeter	+1.07	
Radlova	1939	Photographic photometry	+1.08	
Markov	1948	Photoelectric photometry	—	F9—K2
Teifel'	1962	Spectrophotometry	+0.85	GO

Moonlight is polarized to the greatest extent near the quadratures, when P is as high as 15 or 20%; in the last quarter P is somewhat higher than in the first quarter.

Variations in the polarization of the light reflected by individual lunar details also showed that, on the average, the degree of polarization near the quadratures decreases with an increase in the reflectivity B; therefore, it is higher for "maria" than for "continents" (Figure 104).

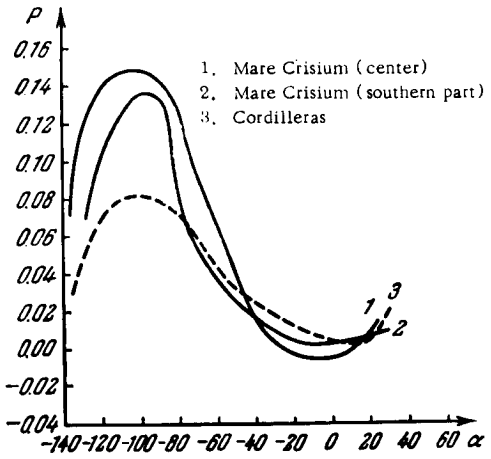


FIGURE 104

However, a recent study /9/ of this phenomenon has shown that near the quadratures the degree of polarization depends not only on the phase angle

α but also on the geological structure of the lunar detail (Figure 105: here α = 79°; the ordinate axis on the right gives the reflectivity; the formation numbers are given according to Fedorets /3/).

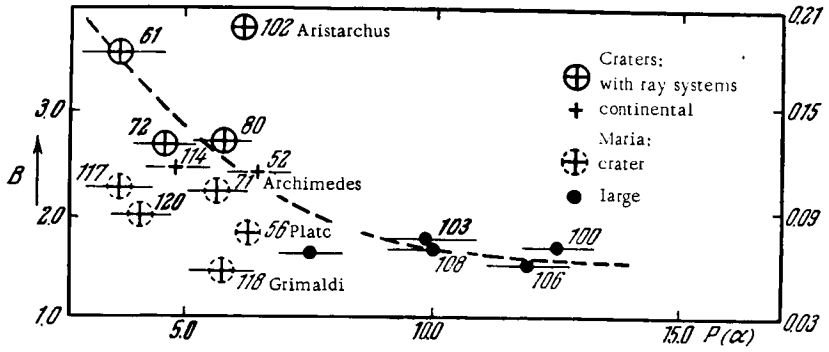


FIGURE 105

It was found that during these phases the surfaces of ordinary and crater "maria" having equal reflectivities differ substantially in polarization and probably in surface structure as well. Crater "maria" [that is, craters lying within a mare*] mostly produce a lower polarization and do not lie on the average $P = f(B)$ curve.

The polarization of the light reflected by lunar formations varies with phase as follows (Figure 104): near new moon P increases, reaching a maximum /10/ shortly before the first quarter at $\alpha \sim -100°$; then P decreases as full moon approaches, until at a phase angle $\alpha \sim -20°$ the polarization disappears. Near full moon (for phase angles from −20° to +20°) polarization appears again, but it differs from the polarization at large phase angles: first P is small, and second the position of the polarization plane is different.

At large phase angles the electric vector of the polarized component of moonlight oscillates perpendicular to the plane passing through the moon, the sun, and the observer (that is, in the plane of the intensity equator); however, near full moon (at phase angles of ±10°), the oscillation plane approaches the plane of the intensity equator. In the former case the degree of polarization of the moonlight is regarded as positive and in the latter case as negative. As regards the appearance of negative polarization at small phase angles, Öhman suggests that it is caused by the mutual illumination of lunar slopes facing one another, which at low phase angles are very inclined to the sun's rays, and by the interaction of the polarized light fluxes scattered from one of these formations to another.

A comparison of the polarization-versus-phase curves for lunar objects and for terrestrial rocks /11, 12, 13, 14/ shows that the lunar curves are close to the polarization curves for volcanic ash. A similar comparison made in /9/ shows that although some lunar areas polarize light just like the ordinary unfused surfaces of meteorites do, still none of the lunar details are capable of producing at quadrature the high degree of polarization found for the fused crusts of meteorites.

* [A more appropriate name would be "marine craters"]

It is also noteworthy that lunar rays are indistinguishable in their polarizing properties from the objects adjacent to them, and that the crater "mare" Archimedes, formed in the midst of the Mare Imbrium, produces the same degree of polarization as the surrounding mare.

§62. **Studies of the temperature of the lunar crust using heat receivers and radio methods.** Another effective method which has increased our knowledge of the structure of the lunar crust considerably is the radiometric measurement of the lunar surface temperatures (see Vol. I, Chap. XXVI, p. 445). The resulting curves showing the distributions of the absolute temperature T and the thermal radiation E over the lunar disk, for different full moons, are given in Figure 106 /6/.

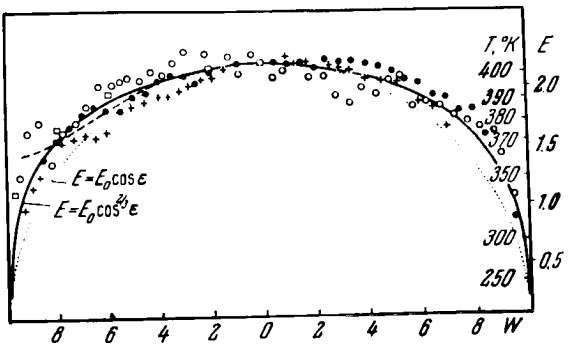


FIGURE 106

The energy per square centimeter emitted by the lunar surface at full moon is given as an average by the formula

$$E = E_0 \cos^{2/3}\varepsilon,$$

where ε is the earthward emission angle, and $E_0 = 2.05 \frac{\text{g} \cdot \text{cal}}{\text{cm}^2 \cdot \text{min}}$ is the energy emitted at the center of the lunar disk. It follows from this formula that the distribution of absolute temperature over the lunar disk should obey the equation

$$T = T_0 \cos^{1/6}\varepsilon,$$

since the absolute emission of the lunar surface is related to the surface temperature by the Stefan-Boltzmann law. In this formula $T_0 = 400°\text{K}$.

The temperature variations of lunar details located at the apparent center of the disk are such that the temperature changes from 390 to 330°K depending on the phase angle and the height of the sun above the horizon (from 90° near full moon to 0° at the quarters); the night side of the moon has been found to have a temperature of about 100 - 120°K.

Such radiometric determinations of planetary temperatures can be made by comparing the total radiation received by a thermocouple, mounted at the focus of a reflector, with the shorter-wavelength component ($\lambda < 1.5\mu$) passing through a water layer 1 cm thick or through a cover glass ($\lambda < 5\mu$).

The long-wave radiation ($\lambda > 8.0\,\mu$) emitted by the sun-heated lunar surface reaches us through a transmission band of the earth's atmosphere lying between 8 and 14μ. The filters mentioned above were used to isolate the radiations emitted by various lunar details during the total lunar eclipses of 1927 and 1939 (Figure 107); these data made it possible to calculate the variations in lunar temperature during rapid changes in the heating of the moon's surface. Observations showed that in 1927 the temperature of a lunar detail which remained in the penumbra for an hour dropped from 340 to 174°K. During the total phase, its temperature dropped from 174 to 156°K in 2 ½ hours. When the detail emerged from the earth's shadow, the temperature rose rapidly from 156 to 339°K.

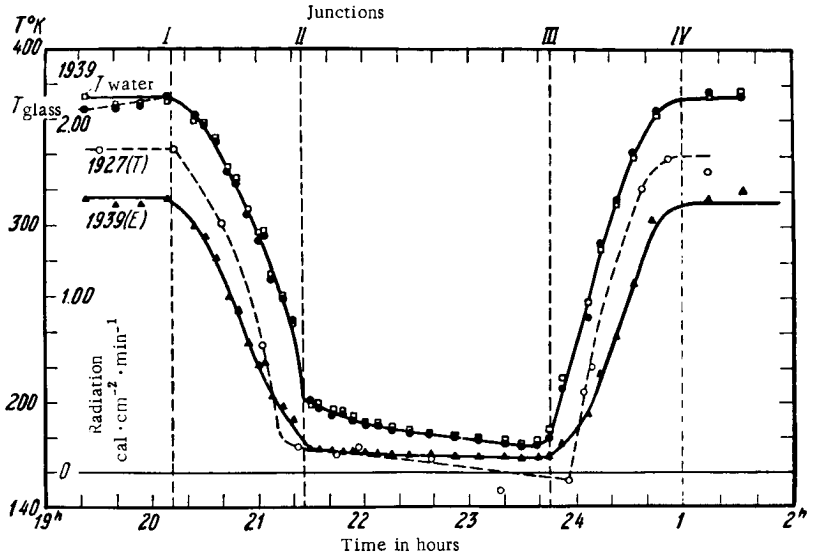


FIGURE 107

The latter, together with the approximately 10-minute lag between the minimum-temperature point and the beginning of the total eclipse, as well as between the onset of the steep temperature rise and the end of the total eclipse, indicate that the material covering the lunar surface has a low heat conductivity. During the total phase of the eclipse the moon received from the sun an amount of energy which was not greater than 0.01 $\frac{g \cdot cal}{cm^2 \cdot min}$, but the energy emitted by the lunar surface dropped only from 0.07 to 0.04 $\frac{g \cdot cal}{cm^2 \cdot min}$.

Observations with thermocouples and filters also made it possible to measure the spherical albedo of the moon for heat radiation; it was found to be 0.13.

In order to compare the average temperatures of narrow strips (3' wide) of the lunar surface, passing over the entire disk of the moon, measurements were made using the Pulkovo radiometer together with thermocouples in a spectral range from 1.4 to 12μ. The modified radiometer of M. S. Zel'tser and A. V. Markov was employed here, and in addition the vacuum

thermocouples specially developed by B. P. Kozyrev. Radiometric measurements made during the full moons on 25 April and 18 September 1956 and 4 April 1958 were processed according to Menzel's modified formula in order to calculate the average temperatures of the strips lying parallel to the celestial meridian at distances of ±0.7 of the lunar radius to the east and to the west of the disk center (Figure 108; curve 1 is for 24/25 April 1956, curve 2 is for 18/19 September 1956, and curve 3 is for 4/5 April 1958), as well as the average infrared albedos of these strips.

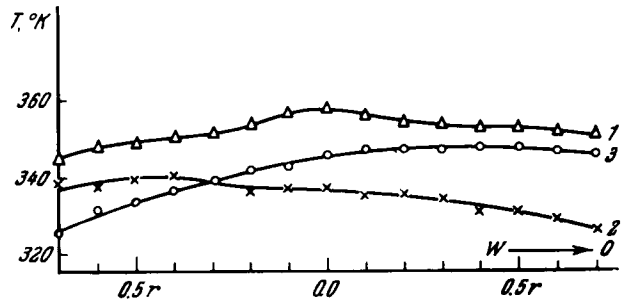


FIGURE 108

The measurements of Markov and Chistyakov show clearly that between phases −15°.5 and +10°.8 the temperatures of the strips depend on the phase angle. The average infrared albedo was found to be 0.12 for a strip which was richer in "maria"and 0.15 for a more "continental" strip.

The processing of the Pulkovo measurements of lunar temperatures necessitated a calculation of the atmospheric absorption of the low-temperature radiation from the moon; this was done by simultaneously determining the water-vapor content /6/.

The average temperature of such a strip, passing during the full moon of 25 April 1956 through the center of the visible disk at a phase angle of 2°.8, was found to be 358°K according to the Pulkovo measurements. Theoretical calculations of the average temperature of this central strip, according to Pettit's measurements of the distribution of T over the lunar disk at full moon, gave an average value of 360°K. Pulkovo measurements of lunar temperatures on different nights agreed to within ±18°, depending on the accuracy with which the atmospheric absorption of the long-wave component could be reckoned. Measurements of these temperatures made during the course of a single good night agreed to within ±4°.

The results obtained so far indicate that such thermoelectric measurements can be used further for determinations of the heat conductivity, in conjunction with radio measurements of the radiation from the moon's crust. At Pulkovo, M. S. Zel'tser /15/ took part in the measurements and calculations of the temperatures of lunar surface zones.

Longer-wavelength radiation, including radio waves, penetrates through a greater thickness of lunar rocks than visual and infrared rays do. Therefore, measurements of the moon's temperature using wavelengths of

$\lambda \sim 1$ mm will give a picture of the variation with phase of the temperatures of lunar rocks lying at greater depths.

Some very important results on lunar temperatures have been obtained with a new radiation detector operating in a wavelength range from 0.2 to 1.6 mm. This optical detector makes use of a Golay infrared cell /16/. Shorter wavelengths (up to $\lambda \sim 0.9$ mm) were cut off with a filter made of black photopaper and vinylite. The upper boundary $\lambda = 1.6$ mm was due to the water-vapor absorption band of the earth's atmosphere, which has its absorption maximum at $\lambda \approx 1.7$ mm.

The detector was mounted at the focus of a searchlight mirror 600 mm in diameter and 240 mm in focal length. It was used to measure the average lunar temperature for different phases, and also the average temperature during the total eclipse of 18 to 19 January, 1954. During the 1954 lunar eclipse, Sinton measured the average temperature of the lunar disk for the entire duration of the eclipse, and the average-temperature curve was found to be less steep than the temperature curve plotted by Pettit according to radio measurements for an individual area on the moon during the 1927 and 1938 eclipses.

When the moon entered the penumbra, the average temperature of the lunar sphere, according to Sinton, was 295°K ± 20°; when it touched the umbra it was 236° ± 10°; at the beginning of the total eclipse it was 180° ± 10°; in the middle of the eclipse it was 175° ± 10°; and at the end it was 170° ± 8°. Finally, at emergence from the umbra and from the penumbra the average lunar temperatures were found to be 175° ± 8° and 180° ± 25°, respectively. A comparison of these data with the average lunar temperature for extra-eclipse days, which according to Sinton is ~350°K in the same wavelength range, and also with the fact that the average-temperature curve of the moon during the eclipse period showed a $1\frac{1}{2}$-hour phase lag relative to the mideclipse point, leads to the conclusion that wavelengths of the order of 1.5 mm give the temperature of the deeper-lying strata of the lunar crust.

During the 1960 eclipse Sinton measured the temperature variations of the floors and surroundings of the craters Copernicus and Tycho as functions of the eclipse phase. The observations showed that the crater-floor temperature (infrared albedo of 0.18) was 50° higher than the temperature of the surrounding areas (infrared albedo of 0.10). These experimental results /16/ are the first to be obtained, throughout the whole history of lunar observations, concerning the dependence of the temperature of a lunar detail on its albedo during an eclipse; however, theoretically such a dependence was to be expected.

Texts on radio astronomy /6, 18/ give determinations of the average lunar temperatures as functions of phase and of radio-wave length in the range from 10 to 12 mm. These texts observe that the average temperature of the moon is found to be lower at higher wavelengths, that is, for deeper-lying lunar strata.

At $\lambda \sim 30$ mm radio measurements fail to detect any phase dependence of the average temperature. The average lunar temperature at such wavelengths was found by Soviet observers to be about 156°K (to within ± 30°).

Australian observers measuring the phase dependence of the average radio-temperature of the moon at $\lambda = 12$ mm have represented their results as follows:

$$\overline{T}_0(\alpha) = 215\ ^\circ\mathrm{K} + 36^\circ \cos(\alpha - 45^\circ).$$

The second term of this formula indicates that at a wavelength of 12 mm the temperature maximum of the respective strata of the moon's crust is reached almost 3 days after full moon /19/.

These phenomena have the following explanation: the deepest strata of the moon's crust, whose radiation is represented by the amplitudes of the 30-mm radio waves, are just barely heated by the sun, and therefore they maintain a constant temperature; the shallower strata, on the other hand, which are represented by the radio waves in the 10-mm range, are actually heated by the sun. The lag* in maximum temperature is analogous to the lag observed in the variation of the temperatures of the deeper strata of the earth s crust during the course of the year.

The actual depths of the corresponding strata of the moon's crust depend, first, on the degree of dispersion of the lunar soil and, second, on the electrical conductivity of the crust. If we assume that the moon's crust has a uniform structure throughout the entire depth from which radio waves are capable of reaching us, then depending on the electrical properties of the material we find that 3-cm radio waves can penetrate through depths from 60 to 9 cm, whereas 1-cm waves can pass through depths from 20 to 3 cm.

The above radio measurements, and also the measurements utilizing the Golay detector and thermocouples, have been used for theoretical calculations of the heat conductivity of the upper layers of the moon's crust /6/. The lunar surface temperature varies periodically with the phase of the moon. Thus, the equation describing the penetration of a thermal wave into the lunar interior, for steady-state conditions and assuming a constant heat-conduction coefficient and also regarding the lunar surface as a halfspace** in which a uniform temperature is maintained over distances many times greater than the penetration depth of the thermal wave, has the form

$$\frac{\partial T}{\partial t} = \frac{k}{\varrho C}\frac{\partial^2 T}{\partial x^2}, \tag{1}$$

where T is the temperature, t is time, k is the heat-conduction coefficient, C is the specific heat, taken as 0.2 for the lunar surface, ϱ is the crust density, taken as 2 near the lunar surface, and x is the depth.

After integrating this equation, we find the following variation of temperature with depth:

$$T = \sum C_0 e^{-x\sqrt{\frac{\pi m}{a^2\tau}}} \cos\left(m\frac{2\pi t}{\tau} - x\sqrt{\frac{\pi m}{a^2\tau}}\right), \tag{2}$$

where τ is the period, m is an arbitrary constant, $a^2 = \frac{k}{\varrho C}$ is the thermal diffusivity, and C_0 is the amplitude for $x = 0$ (at the surface).

At a depth x the amplitude has the value

$$C_0 e^{-x\sqrt{\frac{\pi m}{a^2\tau}}} = C_x, \tag{3}$$

that is, the amplitude decreases with depth. The rate of this decrease increases with the period and the heat-conduction coefficient. The period is independent of depth. If we take the logarithm of (3) and set $m = 1$, then

* Recently this lag was verified by a measurement of the isothermal temperature distribution in the lunar crust; a Soviet radiotelescope and 8-mm waves were used /20/.

** That is, a body limited by a plane on one side and extending to infinity on the opposite side.

we find for the principal wave that

$$\lg^2 \frac{C_0}{C_x} = \frac{x^2 \pi}{a^2 \tau}. \tag{4}$$

At a time t the phase at the surface is

$$\frac{m 2\pi t}{\tau} = Y_0 \tag{5}$$

and at a depth x it is

$$Y_x = \frac{m 2\pi t}{\tau} - x \sqrt{\frac{\pi m}{a^2 \tau}}. \tag{6}$$

Consequently, the phase lag at a depth x relative to the surface is

$$R = Y_0 - Y_x = x \sqrt{\frac{\pi m}{a^2 \tau}}, \tag{7}$$

and for $m = 1$ we have $R^2 = x^2 \frac{\pi}{a^2 \tau}$, indicating that the phase lag increases with depth faster for lower heat-conduction coefficients and periods. As a result, low-period oscillations decay at small depths, so that below this the complex temperature oscillation can be treated as a simple periodic variation correlated with the lunation period. Thus, equation (2) becomes

$$T(x, t) = B + C_0 e^{-x \sqrt{\frac{\pi}{a^2 \tau}}} \cos\left(\frac{2\pi}{\tau} t - x \sqrt{\frac{\pi}{a^2 \tau}}\right), \tag{8}$$

where B is the average temperature.

The radiation from a lunar surface at a temperature T is, according to the Stefan-Boltzmann law, equal to σT^4:

$$\sigma T^4 = A + F_0, \tag{9}$$

where A is the amount of solar energy absorbed per square centimeter of lunar surface per minute, and

$$F_0 = K \frac{\partial T}{\partial x} \tag{10}$$

is the incoming or outgoing heat flux at the lunar surface. The half-amplitude F_0 of the flux at the surface (that is, for $x = 0$) is

$$F_0 = C \left(\frac{2\pi k \varrho c}{\tau}\right)^{1/2}. \tag{11}$$

Wesselink in his calculations proceeds from these equations. In equations (1), (9), and (10) he substitutes for x the quantity

$$\xi = \frac{x}{2\left(\frac{\pi k}{\varrho c} \tau\right)^{1/2}} = \frac{x}{l}. \tag{12}$$

Then, (1) and (10) become

$$\frac{\partial T}{\partial t} = (4\pi \tau \varrho)^{-1} \frac{\partial^2 T}{\partial \xi^2} \tag{13}$$

and

$$F = (4\pi\tau)^{-1/2} (k \varrho c)^{1/2} \frac{\partial T}{\partial \xi}. \tag{14}$$

Wesselink solved equation (14) by numerical integration, and thus he obtained the value of $(k\varrho c)^{1/2}$ both for lunar days and for eclipses. The calculations based on the diurnal variation of temperature gave a value of $0.012 \frac{\text{cal}}{\text{cm}^2 \text{min}^{\frac{1}{2}} \text{deg}}$ for $(k\varrho c)^{1/2}$.

The value of $(k\varrho c)^{1/2}$ calculated by Wesselink on the basis of Pettit's observations of the eclipse on 28 October 1939 was $0.0084 \frac{\text{cal}}{\text{cm}^2 \text{min}}$. This is very close to Epstein's figure $(k\varrho c)^{1/2} = 0.0083$, obtained by the same method for the eclipse of 14 June 1927.

If we insert probable numerical values for ϱ and c ($\varrho \approx 2$ or 3 and $c \approx 0.2$), then we obtain $k \approx 1 \cdot 10^{-4} \frac{\text{cal}}{\text{cm} \cdot \text{min} \cdot \text{deg}}$. This is about three orders of magnitude less than the corresponding value for terrestrial rocks. Therefore, the flowoff of heat into the lunar interior by means of heat conduction constitutes only 0.05 to 0.02 of the incident flux of energy. Measurements show that each square centimeter of the moon's surface radiates from 1.66 to $1.61 \frac{\text{cal}}{\text{cm}^2 \text{min}}$, for normal incidence of the sun's rays. Consequently, according to Stefan's law, the lunar surface temperature is found to be about 377 to 374°K.

All the preceding results seem to indicate the presence of a very porous or dustlike layer on the surface of the moon. This conclusion is apparently corroborated by the laboratory studies of Smoluchowski, who investigated the variation in the heat conductivities of different types of crushed rock as a function of particle size and of the gas content in the particles. The results showed a sharp drop in heat conductivity as the pressure decreased from atmospheric to 0.05 mm Hg.

Various authors have correspondingly obtained values of 120 to $1000 \frac{\text{cm}^2 \text{min}^{\frac{1}{2}} \text{deg}}{\text{cal}}$ for the product $x = (k\varrho c)^{-\frac{1}{2}}$. Therefore, we may conclude that the upper part of the lunar crust is in a crushed (or highly porous) state.

Theoretical calculations lead to two alternative models of the lunar crust. Some theoreticians maintain that the outermost layer (1 to 4 mm thick) consists of finely dispersed, powderlike dust, below which there extends to a great depth a layer of rock debris or pumicelike material. The powdered layer, despite its insignificant thickness, explains the very low heat conductivity of the lunar crust. Other theoreticians do not agree to the existence of a powdered dust layer, and they assume that there is just a fairly deep gravel layer on the lunar surface.

It is still impossible to choose between the two alternatives. Further progress in our knowledge of the physical structure of the lunar surface, however, can be expected as a result of a development of the astrophysical studies outlined above. An exact knowledge of the structure of the outermost layer of the lunar crust, by the way, is essential for an actual landing on the moon to be feasible.

We have observed elsewhere (/16/, Chapters V and VI) that prior to 1960 many studies were made of the temperature, radio emission, and topography of the lunar surface. These results seemed to indicate that the moon is covered with a layer of dust and fine debris several centimeters deep.

At the lunar symposium in Leningrad in December 1960, we supported Firsoff's opinion (/6/ Conclusion) that the outermost layer of the lunar crust may have a low density (about 0.6 g/cm^3). This layer, according to our

calculations, should have a porous structure and may extend to depths up to 0.40 m. This, incidentally, would explain the very low thermal conductivity, and also the photometric and polarizing properties of the light reflected by the moon. Troitskii /21/ read a paper at the Leningrad symposium which described his radio observations of the moon; he observed that, in order to explain the properties of lunar radio emission, the outermost crust layer must have a density of the order of 0.5 g/cm^3 to a depth of 1 to 1.5 m. This result agrees with certain extrapolations to the results of our thermoelectric measurements of the lunar temperatures.

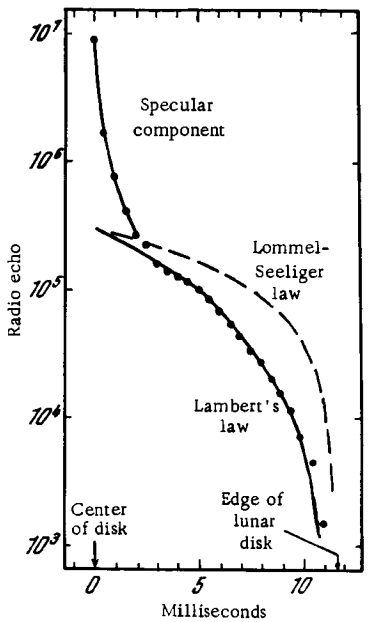


FIGURE 109

However, some areas of the lunar surface (elevated regions such as the ray systems and the floors of craters located in the centers of ray systems) are probably covered by a very thin, low-conductivity dust layer. Sinton offers this as an explanation of the considerably slower drop in temperature during lunar eclipses, a property which distinguishes these ray craters from the rest of the lunar crust.

Minor remains from previously very intense volcanic activity on the moon have recently been detected by Kozyrev /6/, and subsequently by Kalinyak /22/, who studied the spectrum of the floor of Alphonsus. It has been established that in this case gas escapes from a fissure in the floor.

Finally, in preparation for landing on the moon, we must also know the average size of small irregularities and recesses which measure in decimeters.

Published radar data /23/ for various zones of the lunar sphere show that the central part of the disk facing the earth, spanned by a parallel beam of radar rays having an angle of incidence ε from 0 to 29°.8, attenuates the reflected radio beam (for coinciding planes of polarization of the incident and reflected rays) by a factor of 32 (Figure 109). Here, at distances up to 29°.8, the reflected radio beam is scattered by the lunar surface according to the law $\cos^0\varepsilon = 1$, namely just as light is reflected at full moon by the entire lunar surface. It may be assumed that in this case the lunar irregularities are several times greater than the wavelength used, whereas the specular reflection at the center of the lunar disk is possible with irregularities as small as one-twentieth of the wavelength used; the radar waves in this instance were 10 cm in wavelength. It is therefore likely that the fine dust and the small mounds on the horizontal surfaces in the center of the lunar disk facing the earth are no more than 3 - 5 mm in size. At ε = 28.9°, however, such fine deposits are already masked by larger irregularities (of the order of several decimeters). This is also consistent with the fact that, starting at ε = 46°.5 and on up to 80°, an appropriately polarized radar

beam is reflected according to $\cos^1 \varepsilon$ (Lambert's law), which is the same way as visible rays are reflected from a diffuse sphere covered with magnesium-oxide soot. It is known that the soot particles obtained during smoking measure from 0.1 to 1 μ (most of the particles having a diameter of $\sim 0.25\mu$).

We thus have the following situation: if the same dependence (Lambert's law) is also assumed for the reflection of a 70-cm lunar radio wave at $46°.5 < \varepsilon < 80°$, then the average size of the microirregularities in these zones of the lunar globe should range between 7 and 90 cm, the greatest number being concentrated between 25 and 50 cm. All these results will be made more precise by future studies.

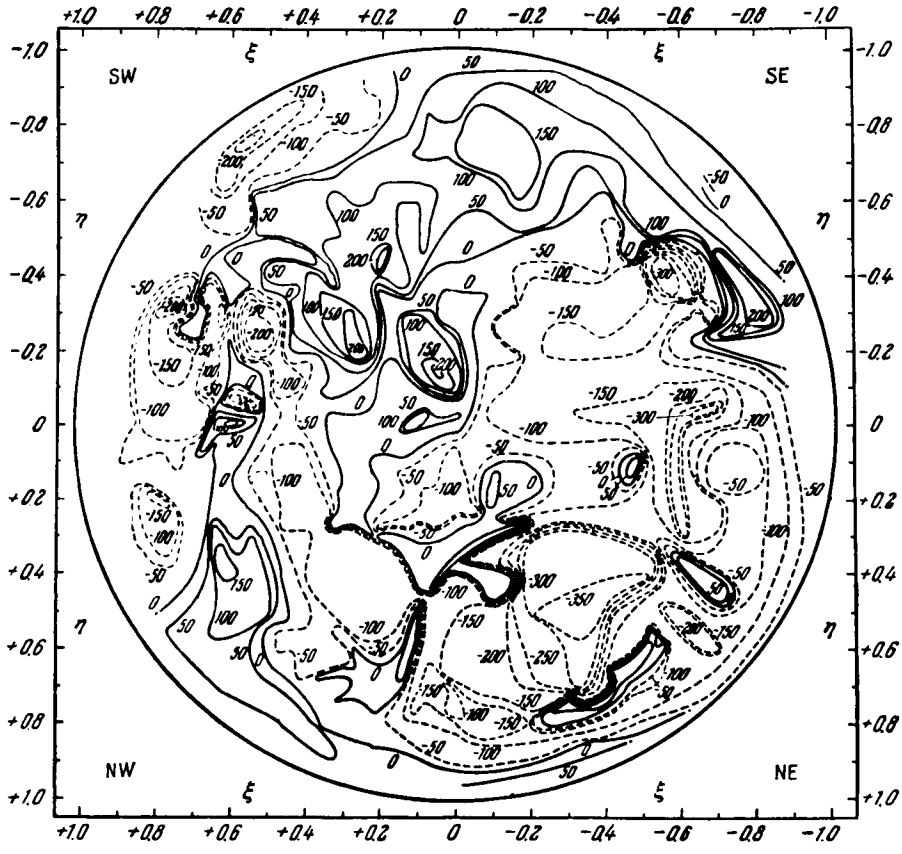


FIGURE 110

With respect to the lunar macrorelief, that is, to specifying the depths of the "maria" in the visible lunar hemisphere, the heights of the mountain ranges, and the elevations of the horizontals at the "maria" and the "continental" shields, the latest stereoscopic-photograph measurements /24/ show that the deepest visible "maria" (Mare Humorum and the eastern part of Mare Imbrium) are deeper by $3.0 \cdot 10^{-3}$ lunar radii (by 5 km) than the mean horizontal of Mare Vaporum near the southern boundary of

the Apennines (Figure 110). The highest plateaus have elevations of 3.5 km above Mare Vaporum. These values represent the most recent data on the nature and structure of the moon's surface.

§63. **Endogenous and exogenous hypotheses of lunar-relief formation.** Several hypotheses have been advanced to explain the origins of the different forms of lunar topography, and they can be divided into two classes. The first class includes hypotheses in which the main role is attributed to endogenous forces, that is, to forces acting from within the moon's interior /6/. The second class, on the other hand, is made up of hypotheses attributing the formation of the lunar topography to external forces, such as the impacts of large meteorites upon the previously solidified lunar crust.

Both groups of hypotheses refer just to the formation of topography after the original solidification of the lunar crust. They do not solve problems related to the origin of the moon as a whole, or related to its primordial state, for instance, the problem of whether the moon originated as a gaseous sphere which gradually became liquefied[magma] and then solidified, or whether it formed as a cold body which subsequently warmed up.

It should be noted that the predominance of circular-symmetric formations on the moon, for example, most of the "maria" and walled-plain craters, shows that they were created by forces acting at their geometrical centers. Advocates of the exogenous theories point out the similarity between the walled-plain craters, which often have a central peak, and terrestrial volcanos, and they maintain that the lunar mountains were formed by phenomena not unlike those of terrestrial volcanic activity. It is argued that the force of gravity on the moon is 1/6 of that on the earth, so that lunar craters might be expected to be considerably larger than terrestrial volcanos.

Advocates of a geologically very fast formation of all parts of the lunar topography, who maintain that the moon necessarily passed through a magma stage and then had an atmosphere during the period of crust solidification, have also tried to take into consideration the possibility of a powerful efflux of gases through the lunar crust, In their opinion, such an efflux would promote vigorous volcanic activity on the moon.

Without going into these assumptions any further, let us now go on to the endogenous hypothesis of lunar-topography formation developed in the USSR in 1949 by Khabakov /14/. Khabakov concluded that the lunar mountains could not have been formed instantaneously (on a geological time scale), and he suggested five "geological" periods for the moon. In his opinion, the oldest lunar formations observable are the "continental" shields, which as a rule have higher reflectivity and are whiter. This original lunar crust was free of circular mountains. Then, like the earth's crust, it subsequently underwent various upheavals and subsidences under the action of endogenous forces.

As a consequence of the first of these crustal expansions, whose traces are still observable, several fissures [rilles] appeared on the moon as well as numerous annular mountains of the Hipparchus type (see Khabakov's geological map of the moon in Reference /6/). These fissures and mountains were spread freely over the entire surface, which during this second lunar "geological" period did not contain a single "marine" lowland.

The state of the lunar crust preceding the expansion during the second period is represented by the almost virginal craterless areas of the Apennines, the coasts of Sinus Iridum, Mare Crisium, etc.

During the third period the lunar crust, then covered in many places by primeval walled-plain craters, subsided under the action of internal forces. This gave rise to some of the oldest "maria," the only one to emerge undestroyed from the subsequent orogenic periods being the enormous "mare" which occupied the place of the present-day Mare Nectaris; originally, however, this "mare" extended to the Altai range, which then served as its "coast."

A new expansion of the lunar crust, following this "Altaic" period, led to the fourth geological period on the moon. Then, simultaneously with the appearance of several new fissures, a great number of walled-plain craters were formed, including Theophilus, Cyrillus, Catharina (partly tapering off the "Altaic" lowland), Ptolemaeus, Schickard, Plato, and also Fracastor and Fra Mauro. The last two craters originally formed on a "continent," which Mare Nectaris and Oceanus Procellarum were subsequently to replace; the formation of the latter naturally modified the shape of the craters.

The belt of "younger" lunar "maria," among which Khabakov includes Mare Crisium, Mare Foecunditatis, Mare Nectaris, Mare Tranquillitatis, Mare Serenitatis, Mare Imbrium, and Oceanus Procellarum, were caused, according to Khabakov's hypothesis, by a final subsidence of the lunar crust, which occurred during the fifth "geological" period. It was then that magma flowing into the fissures in the new "maria" partly destroyed the walls of Sinus Iridum in Mare Imbrium and the walls of the crater Fracastor in Mare Nectaris and also almost entirely obliterated the walls of Fra Mauro.

These upheavals and subsidences of the lunar crust mark the gradual dying out of orogenic activity on the moon. The last crust expansion, which occurred in various areas during the sixth "geological" period (the most recent one), produced 300 young craters of the type of Tycho and Copernicus. The eruptions of these craters, which are now distinguished by high albedos, gave rise to the systems of bright rays which we discussed above.

However, this view of the origin of the rays is by no means that of all selenologists, since some rays do not issue from the center of the parent crater but rather lie tangent to the crater walls. Some selenologists maintain that, during the expansion of the lunar crust which led to the formation of the ray systems, radial fissuring took place, the parent craters being created at the centers of the fissure systems. According to some hypotheses, the radial fissures and the necks of the central craters of these ray systems were filled up with lava from the moon's interior; the solidified lava thus accounts for the high albedo.

In the present epoch (more precisely, during the last 100 years of regular observations of the lunar surface) the orogenic activity of the moon has temporarily abated.

It should be noted that volcanic activity on the moon in 1956 - 1958 would explain the very interesting changes in certain features on the floor of Alphonsus. In the western part of Alphonsus there is a fissure which does not always look the same, even under identical conditions of solar illumination, and which has been recorded differently in such cases on ordinary and infrared photographs of the crater floor. In some cases the fissure was easily observable through a telescope eyepiece and on blue and infrared plates. In other cases, however, with analogous illumination, it was either not visible or else poorly visible (both visually and in blue light). This

phenomenon led the American astronomer Alter to suspect that a transient efflux of gases from the moon's interior took place through this fissure on 26 October 1956.

The intermittent escape of gases in Alphonsus was finally established by a unique spectrogram of the central peak. This spectrogram was taken by Kozyrev on the night between 2 and 3 November 1958, using the spectrograph of the 50" telescope of the Crimean Observatory /6/.

It should be mentioned,, too, that in October 1958 Kozyrev had obtained some spectra of other areas in this crater, near the fissures whose appearances in red and blue light had been studied by Alter in October 1956. However, prior to 2 November 1958 Kozyrev detected nothing out of the ordinary in his spectra of the Alphonsus floor. Then, on the night between 2 and 3 November 1958, Kozyrev, aided by Ezerskii, obtained several spectra of the central peak of Alphonsus, with the setting sun not high above the floor and with a phase angle $\alpha = +77°.6$, these being almost the same conditions as those under which Alter obtained his photographs of Alphonsus in 1956.

When the Soviet observers, who also photographed on the same night some other spectra of the Alphonsus floor (including the region of the fissures which interested Alter), examined their results, they found /6/ a gas-emission spectrum in the photograph of the central peak of Alphonsus obtained on 3 November 1958 between 3^h and 3^h30^m UT. It should be noted that another spectrum of the same peak, photographed between 3^h30^m and 3^h40^m UT, showed no traces of this emission. This was regarded by the observers as proof of the transiency of the phenomenon.

Since Kozyrev's spectra were all calibrated, the nature of the phenomenon which occurred in the crater on the night between 2 and 3 November 1958 could be ascertained. This was done by means of a photometric processing of the spectra, carried out at Pulkovo by Kozyrev and Kalinyak.

Aside from the action in Alphonsus, only three comparatively insignificant changes on the moon have been reported in all the relevant literature:

1) Linné, a crater in Mare Serenitatis, shrank from 10 km in diameter prior to 1866 to not more than 4 km after 1866;

2) a new crater about 4 km in diameter appeared in the neighborhood of the rille in Hyginus (near the center of the lunar disk);

3) the reflectivity of some areas on the floor of the walled plain of Plato changed in 1870 - 1923.

With regard to exogenous orogenic hypotheses, it should be mentioned that Gruituisen (1825) was the first to cite the impacts of meteorites upon the exposed lunar surface, not protected by an atmosphere, as possible causes of the lunar topography. Calculations have shown that in such a case, with meteorite velocities of about 40 km/sec and adequate masses, an impact should result in several explosions sufficient to form craterlike cavities.

The collision hypothesis (as opposed to all the others) does not require a molten lunar crust, since "explosion" craters may form in a solid crust. This happens occasionally on the earth as well, and four terrestrial meteoritic craters of recent origin have been found. However, if typical lunar features are examined with the object of verifying this hypothesis, it very soon becomes clear that external forces cannot be regarded as the main orogenic factor.

Actually, the following points are true:

1) due to the orbital motion of the moon, most such craters should form near the lunar equator, whereas in fact they prevail near the poles;

2) the mean frequency of small lunar craters on the "maria" and by the walls of larger craters should be the same according to this hypothesis, whereas in fact small craters are more numerous by the large-crater walls than on the "maria";

3) a number of authors have established the following relation: the younger the lunar crater is, the smaller is its diameter and the lower are its walls. Clearly, the fall of large meteorites could not have been confined just to the distant past.

Finally, the collision hypothesis cannot explain the creation of the lunar "maria," although some authors /6/ believe that it can. Therefore, it seems more likely that the main forces responsible for the formation of the lunar topography were internal in origin. It is quite probable at the same time that a small number of the lunar craters not more than 5 or 10 km in diameter are collision craters produced by large meteors.

Since at this time considerable effort is being concentrated not only on problems of cosmogony but also on studies of the structure of the surface layer of the lunar crust with a view toward achieving the landing of spacecraft on the moon /25/, therefore astrophysics has been able to contribute greatly to the knowledge of this structure. Let us now summarize the situation as described in the preceding.

1. The entire lunar surface is highly pitted. In places it is covered by rilles, some large enough to be telescopically visible and some not. Most of the surface irregularities are less than 40 m in size and cannot be seen through a telescope.

2. The lunar surface has a low reflectivity, decreasing toward the ultraviolet. Near $\lambda\lambda$ 390 and 480 mμ local drops on the curve are observed, possibly indicative of the presence of SiO_2 on the moon.

3. The polarization of the light reflected by the lunar surface is close to the polarization of light reflected by volcanic ashes. The degree of polarization is much lower than that for light reflected by the fused crusts of meteorites.

4. Studies of the heat conductivity and electrical conductivity of the lunar surface suggest the following two models for the lunar-crust structure:

a) the moon is covered by fine gravel from 10 to 60 cm deep, or else it consists of porous rock to this depth /21/;

b) the existing thermal-conduction and electrical-conduction properties of the lunar crust may be reproduced in terms of a two-layer model, the upper layer having a thickness of no more than 2 to 4 mm and consisting of powderlike fine dust; the inner layer may consist of pumicelike rock.

The following three hypotheses have been advanced concerning the structure of the lunar crust.

I. Some astronomers and geologists /6/ maintain that the primary rock juts out to the surface in places as cliffs (for example, on mountain ranges).

II. Advocates of the collision theory /4, 6/ maintain either that the moon is covered by about a centimeter of meteoritic dust, the debris of the explosions of meteorites hitting the surface of the moon, or else that these explosions vaporized not only the meteorites but also part of the crustal matter near the impact point. If the latter was the case, then in their

opinion the vaporized material would have solidified and settled onto the lunar surface as a high-temperature modification of the original substance.

Since the lunar crust is apparently similar to igneous rocks on the earth, therefore the fused or slaglike product should be dark in color (like the fused crust of meteorites which have passed through the earth's atmosphere).

III. Advocates of the endogenous hypothesis of lunar-topography formation believe that the surface of the moon may to a large degree be covered with volcanic ashes.

While choosing between these hypotheses, we should remember first of all that the lunar surface definitely does not possess the homogeneous topography which follows from the premises of the collision theory. Furthermore, a final solution of the problem of the structure of the moon's surface (both the microrelief and macrorelief) is impossible without taking into account the latest Soviet /20/ and American /23/ radio studies of the moon.

§64. Maps of the other side of the moon. The present chapter was written at a time when the accelerating progress of Soviet science had already enabled the USSR to launch toward the moon three rockets, the second of which actually hit the lunar surface.

TASS reports show that the first Soviet lunar rocket, launched on 2 January 1959, passed at a distance of 6000 km from the moon and then became the first artificial planet revolving about the sun. The flight of this rocket near the moon gave the first indication that the earth's satellite has virtually no magnetic field. This rocket also confirmed that there will be little danger from meteors during space flights /25/.

The second lunar rocket, launched in the USSR on 12 September 1959, reached the moon and crashed on its surface on 14 September at $0^h2^m24^s$ Moscow time. Radio determinations of its trajectory at the moment of nearing the moon made it possible to calculate to within ±100 km the selenographic coordinates of the landing point. According to these data it landed in Mare Imbrium, in a region near the eastern wall of Autolycus with coordinates $\varphi = 30° \pm 1°.8$ and $\lambda = 0° \pm 3°.6$.

This rocket also confirmed the almost total absence of a magnetic field around the moon, and it established that the artificial radioactivity of the lunar surface (which is unprotected by an atmosphere), a radioactivity produced by short-wave and X-ray solar radiations and by cosmic rays, is much below the expected value. Moreover, the rocket instruments recorded the possible presence of a weak accumulation of cosmic particles at a distance of about 10,000 km from the moon.

The third Soviet lunar rocket, launched on 4 October 1959, put the first automatic interplanetary station (AIS) into orbit. The AIS took the rocket around the moon, so that from a distance of about 60,000 km it could photograph the side of the moon which is invisible to terrestrial observers. The USSR Academy of Sciences reports /26, 27/ show that these photographs of the other side of the moon were made when the automatic station was over an area with selenographic coordinates of +17°N and −120°W. This corresponds to the latitude of Mare Crisium, about 60° west of the latter.

The published photographs of the other side of the moon have very little in common with previous hypothetical charts of it. The photographs show that the other side of the moon is mainly "continental," and that in its northern part are observed only two new "maria," together with many craters.

The zone between −110°W and −80°W in this hemisphere apparently has a lower albedo than the rest does, but its reflectivity is close to that of the "pali" on the visible side. Thus, it follows that Wilkins was mistaken when he suggested that in the center of this hemisphere there is a "marine" lowland of the type of Oceanus Procellarum. Selenologists may consequently have to modify their previous views concerning the figure of the whole moon as a celestial body.

Soviet scientists have published several works /28-32/ dealing with the processing of the photographs of the other side of the moon taken by the third Soviet rocket on 7 October 1959; these studies also discuss the structure of the invisible lunar hemisphere.

Let us first consider the latest data on the structure of the averted side of the moon. An attempt to arrange this data into a coherent system was made at Pulkovo Observatory, where a special map of the other side of the moon was drawn up (Figure 111) /29/, during the course of preparatory work on "The Atlas of the Other Side of the Moon" /27/, subsequently published by the USSR Academy of Sciences.

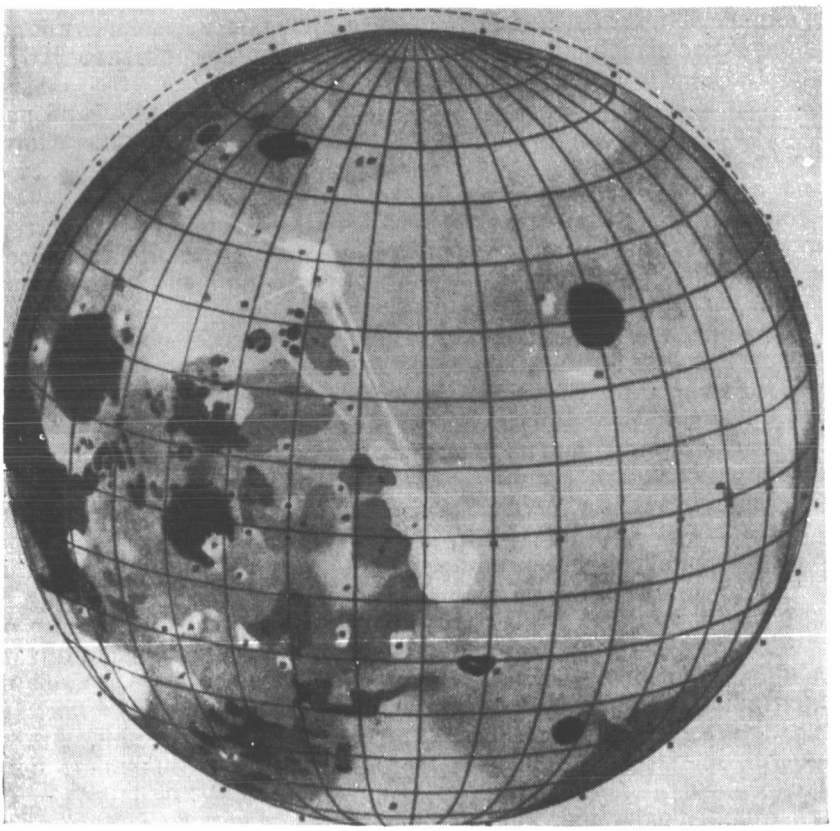


FIGURE 111

When the Soviet map and catalog of the other side of the moon had been completed, the reliability of the conclusions drawn was verified /33/ by an

analogous treatment of photographs of the visible side of the moon taken at Pulkovo during the full moon on 13 March 1960. These Pulkovo photographs were also reproduced in a laboratory using the AIS equipment (to simulate the original photographs of the other side), and then the televised images were processed at Pulkovo. A preliminary study was next made of the accuracy of the photometric analysis of the surface details of the other side of the moon /30/, using only the televised images of the original photographs. The photometric analysis of the televised photographs of the other side of the moon led to the following conclusions concerning the surface details there.

1. The "maria" (Marginis, Smythii, Moscovianum) and parts of "maria" (Humboldtianum, Australe) lying on the other side of the moon represent the same type of "marine" lowland as can be seen on this side of the moon. New spots similar to lunar craters with darker floors are observed near Maria Marginis and Australe. The Sea of Hope marked on maps of the other side is brighter than the lunar "pali" or, at any rate, is no less bright than the "pali."

2. The floor of the crater Tsiolkovskii appears to be darker than the darkest floors of other craters and maria.

3. The large area with brighter dark-gray and light-gray regions, situated on the other side of the moon between Mare Smithii and the Soviet Mountains, is comparable in reflectivity to the lunar "continents" of medium brightness with "pali" and half-submerged walled-plain craters of the type of Ptolemaeus and Petavius.

4. The light part of the "continent" beyond Mare Moscovianum and lying near the north pole on the other side of the moon is comparable in brightness to the light-colored "continental" shield which on the visible side surrounds Tycho and is covered with a great number of annular mountains.

5. The crater Giordano Bruno and the Soviet-Mountain region on the other side of the moon are comparable in reflectivity to the centers of the ray systems of Tycho and Copernicus on the visible side. This also ties in with the presence of individual rays extending from these formations. We have concluded that, within the margin of error of our photometric analysis, the surface details on the other side of the moon apparently do not differ from similar formations on this side in either their degree of surface unevenness or their range of reflectivity gradations.

The above conclusions were tested by means of photometric and other analyses of the Pulkovo photographs of the visible hemisphere of the moon taken during the full moon of 13 March 1960 /33/. Without going into a detailed discussion of this work, let us note that the photometric analysis of the televised images of the visible hemisphere (these images having been obtained by the same methods as those used for the photographs of the other side of the moon taken on 7 October 1959) enabled us to calculate the known reflectivities of various details on the visible side of the moon to an accuracy of about ±20%. The Pulkovo negatives, when televised using the AIS equipment and then processed, provided us with certain more or less definite criteria concerning the identification of known crater types on the other side of the moon (for conditions of illumination close to full moon). Thus, we were able to draw the following conclusions.

1. By means of the photometric analysis and examination of such negatives it is difficult to detect the primeval, frequently huge, craters of the first orogenic period, marked yellow on Khabakov's geological map (/6/, Chapter

VII, p. 241). Under the given conditions the brightness distribution over the surface cover of these walled-plain craters is independent of their geological structure; it is masked by later processes occurring in the surface layer of the lunar crust.

2. Craters from the second orogenic period of the moon (Khabakov's system), and particularly craters in "maria" (of the type of Pitatus and Ptolemaeus), are detectable due to their somewhat lower floor albedos. However, the televised photographs showed only about 40% of the actual number of these formations appearing on the original Pulkovo photograph of the moon taken on 13 March 1960.

3. The young walled-plain and other craters of the third orogenic period (Khabakov's system) always show up on the televised photographs, due to the high brightness of their walls and halos (these craters, of the type of Tycho, are surrounded by ray systems). Consequently, for future investigations of the other side of the moon, we must realize that all the lunar details transferred onto the Pulkovo map using the televised negatives, and included in the first category in "The Atlas of the Other Side of the Moon" /27/, actually exist. The nature and authenticity of the formation known as the Sea of Hope, however, constitute a problem to be solved by later studies of the other side of the moon.

In the future, a more detailed investigation of the invisible lunar hemisphere will enable us to compile a map more accurate than the one in Figure 111; also, it will lead us to more correct conclusions concerning the surface of this hemisphere, some 80% of which has been photographed. Since the photographs also show an area adjoining Mare Crisium and Mare Foecunditatis in which regions of known albedo have been identified, the approximate photometric analysis of these photographs leads to the following conclusions concerning the nature of the formations on the other side of the moon: several bright centers of ray systems are observed, and also an extensive bright "continental" zone with individual walled-plain craters hundreds of kilometers in diameter and individual dark-floored craters. The bright "continental" areas on the other side of the moon are comparable in reflectivity to the "continent" occupying the southern part of the visible hemisphere.

The albedos of the "maria" (Moscovianum, Marginis, and Smythii) are close to those of the dark "maria" on the visible side. The part of the other side of the moon between the Soviet Mountains and Mare Crisium has a high albedo, corresponding to those of the mountainous areas on this side of the moon known as "pali." Certain dark-floored craters on the other side of the moon, including Tsiolkovskii, are the darkest lunar features known; their albedos are close to that of the floor of Plato.

Finally, the smoothness factor of the other side of the moon and its brightness distribution during full moon apparently do not differ from the values on this side. This leads to the important conclusion that both lunar hemispheres have the same surface structure.

Some new data for studies of the lunar region and of the moon's surface were obtained by the fourth Soviet lunar rocket, launched on 2 April 1963. According to the TASS newspaper reports on 3, 4, and 7 April, the last stage of the rocket was first put into an intermediate orbit around the earth, from which it subsequently took off to move into its specified trajectory. The payload of this rocket weighed 1422 kg, comprising the automatic station Luna IV. The object of this flight was to study the moon's surface from

very close, using the automatic equipment installed in Luna IV; the corresponding data on the structure and physical conditions of the moon are important in the preparation for landing astronauts there.

At 1^h24^m UT on 6 April 1963, Luna IV passed at a distance of 8500 km from the moon. All the experiments and measurements using Luna IV were completed according to schedule, and radio communication between the earth and the capsule was maintained for some time after this. The experimental material obtained will be very helpful in solving the technical problems related to conquering the moon. The processing of this material will soon be completed and published.

REFERENCES FOR CHAPTER X

1. BARABASHOV, N.P. — Astronomische Nachrichten, Vol.217. 1923.
2. MARKOV, A.V. — Astronomische Nachrichten, 221: 55. 1924.
3. FEDORETS, V.A. — Trudy Astronomicheskoi observatorii Khar'kovskogo gosudarstvennogo universiteta, Vol.2(10). 1952.
4. SYTINSKAYA, N.N. Luna i ee nablyudeniya (The Moon and Its Observations). — Gostekhizdat. 1956.
5. STAIR, R. and R.JOHNSTON. — J.Res. Nat. Bur. Stand. (USA), 51:81. 1953.
6. Luna (The Moon), Edited by A.V.Markov, Chap.V, §2, §3. — Fizmatgiz. 1960.
7. TEIFEL', V.G. — Trudy Sektora astrobotaniki Akademii nauk Kazakhskoi SSR, 8:131. 1960.
8. SHARONOV, V.V. — Uspekhi astronomicheskikh nauk, 6:181; Izdatel'stvo Akademii nauk SSSR. 1954.
9. MARKOV, A.V. — Izvestiya Glavnoi astronomicheskoi observatorii, No.158: 138. 1958.
10. KOKHAN, E.K. — Izvestiya Komissii po fizike planet, No.1:41. Kharkov. 1959.
11. LYOT, B.—Comptes Rendus, 177: 1615. 1923.
12. LIPSKII, Yu.N. — Trudy Gosudarstvennogo astronomicheskogo instituta im. Shternberga, 22: 66.1953.
13. BARABASHOV, N.P. — Astronomische Nachrichten, 229 (5731). 1927.
14. KHABAKOV, A.V. Ob osnovnykh voprosakh istorii razvitiya poverkhnosti Luny (On the Basic Problems in the History of the Development of the Lunar Surface). — Geografgiz. 1949.
15. MARKOV, A.V. and Yu.N.CHISTYAKOV. — Izvestiya Glavnoi astronomicheskoi observatorii, 21(163):166. 1960.
16. SINTON, W.M.—Journ. Opt. Soc. of Amer. 45(11): 975. 1955; SINTON, W.M. — Ap.J.123(2): 325. 1956.
17. SINTON, W.M. — Sky and Telescope, 19(6):346. 1960.
18. Trudy V soveshchaniya po voprosam kosmogonii (Proceedings of the Fifth Conference on Problems of Cosmogony), pp.325-335. — Izdatel'stvo Akademii nauk SSSR. 1956.
19. PIDDINGTON, J.H. — Aust. J. Sci. Res., 459. 1951.
20. AMENITSKII, N.A., R.I.NOSKOVA, and A.E.SALOMONOVICH. — Astronomicheskii zhurnal, 37(1). 1960.
21. TROITSKII, V.S. — Izvestiya Komissii po fizike planet, No.3: 16. 1961.
22. KALINYAK, A.A. — Trudy Leningradskogo simpoziuma MAS* po Lune v dekabre 1960. 1963.
23. GREEN, P.E. and G.H.PETTENGILL. — Sky and Telescope, 20(1). 1960.
24. BALDWIN, R.B. — Sky and Telescope, 21(2). 1961.
25. ALEKSANDROV, S.A. and R.E.FEDOROV. Sovetskie iskusstvennye sputniki i kosmicheskie korabli (Soviet Artificial Satellites and Space Ships). — Moskva. 1961.
26. Pervye fotografii obratnoi storony Luny (The First Photographs of the Other Side of the Moon). — Izdatel'stvo Akademii nauk SSSR. 1959.
27. Atlas obratnoi storony Luny (Atlas of the Other Side of the Moon). — Izdatel'stvo Akademii nauk SSSR. 1960.
28. MIKHAILOV, A.A. — Vestnik Akademii nauk SSSR, No.1:39. 1961.
29. BREIDO, I.I., A.A.MARKELOVA, and D.E.SHCHEGOLEV. — Iskusstvennye sputniki Zemli, No.9:30.1961.
30. MARKOV, A.V. and D.E.SHCHEGOLEV. — Iskusstvennye sputniki Zemli, No.9:49. 1961.
31. KHABAKOV, A.V. — Iskusstvennye sputniki Zemli, No.9: 53. 1961.
32. LIPSKII, Yu.N. — Sky and Telescope, Vol.21. March 1961.
33. MARKOV, A.V. — Iskusstvennye sputniki Zemli, No.9: 141. 1961.

* [Mezhdunarodnyi astronomicheskii soyuz (International Astronomical Union).]

V. V. SHARONOV

N 67-15495

Chapter XI

PHYSICS OF THE PLANETS

§65. Introduction. The study of the planets and satellites in the solar system is difficult for the following reason. Due to the low temperatures of these celestial bodies, self-radiation at frequencies corresponding to the ultraviolet, visible, and near-infrared regions of the spectrum does not occur. Consequently, these bodies are observed predominantly in reflected sunlight. This means that the basic tool of astrophysical research, spectroscopic studies, can only to a limited extent be applied to planetary observations. Actually, here spectroscopy can be used fruitfully only to study the chemical compositions of planetary atmospheres, on the basis of the absorption spectra obtained. With respect to atmospheric aerosols and the solid surfaces of the planets and satellites, on the other hand, the spectroscopic method gives negligible results, since reflection or scattering of the sun's rays hardly ever produce characteristic lines and bands in the planetary spectrum. Thus, the physical research in this case is based on photometric measurements of various kinds, which are then analyzed to obtain the absolute reflectivities for different spectral regions, the relative angular distribution of the reflected flux, and also data on the polarization and spectral composition of the reflected light.

A physical interpretation of these data is quite difficult. The proper study of light scattered in the gases of atmospheres and cloud systems became possible only during the last decades, thanks to the considerable progress made in scattering theory (by Academician Ambartsumyan in the USSR, by Chandrasekhar in the USA, and by others). However, even now this study just gives certain data of physical nature and tells us nothing about the chemical composition.

The situation is even less favorable in connection with the solid surfaces of planets and satellites. The colors of minerals, rocks, and various types of weathering crusts are extremely variable, so that it is difficult to draw any conclusions concerning the nature of the surface from its reflectivity. Furthermore, a very thin layer of extraneous matter covering the surface proper may completely change the color of the apparent surface. Thus, no theoretical criteria exist for the interpretation of photometric and colorimetric data. Analyses in this field are therefore restricted to purely hypothetical arguments based on analogies with the colors of terrestrial materials and landscapes.

Somewhat more reliable are the results obtained by radiometric measurements in the far infrared, that is, at frequencies where the planetary heat is quite intensive. A proper interpretation of these data gives the surface temperatures of the planets and satellites. Consequently, further observations of the planets at radio frequencies should be very rewarding, since

radio studies have already yielded valuable information on the nature of the moon and the planets.

It follows from the foregoing that the classical information obtained by astronomical observations and celestial-mechanical computations still constitutes the most valuable and reliable data available to the planetologist today. Such classical data are the sizes, figures, masses, and mean densities of the planets, their rotation elements, and also certain cartographic and topographic material.

§66. **The planetary disk.** When observed through a quite powerful telescope, a planet no longer appears to be a starlike point in the sky; rather, the planetary sphere and various surface details become distinguishable. The planets are observed in projection onto the celestial sphere. The region of the celestial sphere occupied by the projection of the body of the planet is called the planetary disk.

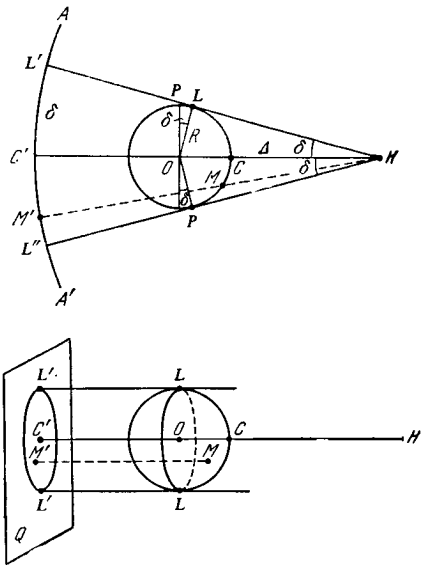


FIGURE 112

An observer views the visible hemisphere of a planet in external perspective projection (Figure 112, top). This means that any point M on the surface of the planet is projected out to a point M' on the celestial sphere AA, at which point the sphere intercepts the line HM drawn from the observation point H to the given planetary detail. The disk is delimited by the line of intersection between the celestial sphere and the cone $L'HL''$ tangent to the surface of the planet. This line is known as the limb of the disk. The point C', which is the projection of the body center O, as well as the surface point C on the line OH, will obviously be the center of the disk. For a spheroidal celestial body, the disk shows up as a small circle whose angular radius δ is easily calculated from the right triangle OLH, where $OL = R$ is the linear radius of the body and $OH = \Delta$ is the distance of the body from the observation point (in practice, the

geocentric distance):

$$\sin\delta = \frac{R}{\Delta} \tag{1}$$

or, since δ is always small,

$$\delta'' = \frac{R}{\Delta \sin 1''} . \tag{2}$$

This formula is used in calculations of the planetary radius R from the measured angular radius δ, or, conversely, in calculations of the angular radius ("semidiameter") from known values of R and Δ. Values of δ are given in astronomical almanacs.

The drawing shows clearly that the radius OL drawn through the tangent point does not coincide with the perpendicular OP to the line OH. Thus, it follows that on the surface of the planet the limb is a small circle perpendicular to OH and displaced relative to its copolar great circle by an angle $\angle POL = \angle LHO = \delta$ toward the observer. Consequently, an observer in general sees less than the hemisphere of the planet.

Since for planets δ is not greater than 35", while even for the moon it is close to 16', the angle POL can be neglected and the limb can be regarded as the great circle normal to the line OH. Moreover, the portion $L'L''$ of the celestial sphere can be regarded as a plane, so that the external perspective projection becomes the simpler, orthographic projection shown in the bottom drawing of Figure 112.

Here Q is the plane perpendicular to OH onto which the image of the planet is projected; it is called the plane of the figure [or image plane]. Since the angle δ has been neglected, all the projection lines are now parallel to the projection axis OH, and any surface point M will be projected to the base point M' of the perpendicular MM' drawn from the point in question to the plane Q.

§67. Methods of observing planetary features. The main technique used to study the surface details of bodies in the solar system is still just the visual observation of these features through a powerful telescope with the highest possible resolving power, A necessary condition for the success of such observations is calmness of the earth's atmosphere. Consequently, planetary observations should be staged in regions with the best visibility, and each observation should cover a considerable period of time, so that the observer can obtain as many sharp and undisturbed images as possible. Records of the observation are generally drawn up by hand while the observer is at the telescope, and a detailed description is noted down. The drawings thus obtained are then used to make a map of the planet. In some cases the positions of the sharpest features are measured with a micrometer, but experience shows that this gives only a negligible gain in accuracy.

For a photographic investigation of the surface features of a planet, the planetary disk must be photographed on the largest scale possible. Thus, planets should be photographed using long-focus instruments, which have equivalent focal lengths of about 10 to 20 m. When an astrograph with a shorter focal length is used, additional magnifying equipment must be provided.

Atmospheric vibration has even more influence on photographic observations than on visual ones, because photographic equipment cannot "catch" a moment of quiet image. Furthermore, since the photographic materials

used in astronomy have a relatively low resolving power, photography is inferior to visual observations whenever features of small angular size are involved.

On the other hand, photographic methods can detect low-contrast features which are imperceptible to the eye. This is possible through the application of photographic and printing techniques which increase the contrast of the image. The principal contrast-increasing technique involves printing the positive not from one but from several negatives, successively projected onto the same spot of the positive; in this way complex images with heightened contrast of real details are obtained. Such a multilayer technique also lowers the contrast of emulsion grains and of random defects.

Another useful technique is filter photography in narrow spectral regions. The efficiency of such methods was shown by Tikhov back in 1909. Many valuable results were subsequently obtained by photographing the various planets in ultraviolet light. For example, ultraviolet photographs detected the "violet" clouds on Mars, they established the existence of belts and located the position of the equator on the disk of Venus, etc.

Still further progress in obtaining large-scale, high-resolution images of the planets is to be expected with the introduction of electron-optical converters and multipliers, and also of television techniques, on the one hand, and with the beginning of extra-atmospheric observations by astronauts on the other.

§68. Methods of disk measurement. It is very important to measure accurately the angular diameters of the disks of planets and satellites, since then, using formula (2), it is possible to calculate the linear semidiameter and the oblateness of the celestial body. The following methods are used for such measurements.

1. Transit instruments. The angular semidiameter is here obtained by observing the transit of the disk limb through a given meridian; this is done for the moon and the major planets. The usefulness of the transit method is limited, since it only gives the semidiameter coinciding with the diurnal parallel, and additional phase corrections must thus be introduced.

2. Cross-hair micrometer. In this case the micrometer cross hairs can be centered on the limb for any position angle. This method is widely used, but the results are burdened by systematic errors due to irradiation and due to the particular way in which the cross hairs are placed tangent to the limb ("the contact error"). The disk diameter obtained is generally too large.

3. Heliometer. This is an astrometric instrument comprising a refractor whose objective is bisected along its diameter, so that the two halves can move (one relative to the other) along the division line. This displacement doubles the image, and when the two images are placed so that they touch at the opposite edges the contact error is reduced somewhat. Nevertheless, the systematic errors are appreciable, the measured diameter being generally less than the actual figure.

4. Double-image micrometer. Different variants of this instrument all utilize the same principle, that of bringing into contact two images of the disk. However, here the optical device is placed at the eyepiece node of the equipment. This constitutes a certain advantage over the heliometer, since the instrument can be mounted on any telescope, including large

refractor and reflector types, so that observations can be made at higher resolving powers.

5. Disk micrometer. Here an artificial disk is produced in the telescopic field of vision, the apparent disk size being variable within certain limits. The brightness of this disk is first made equal to the brightness of the observed object, and then the apparent dimensions are equalized. Unlike the previous methods, in which the absolute random error is constant so that the relative error increases in inverse proportion to the angular diameter of the measured disk, the disk micrometer has a constant relative error, and this makes the instrument very useful for measurements of disks with small angular diameters.

6. Photographic measurements. This method is the least reliable, because photographic irradiation always blurs the limb of the image, whereas the linear diameter of the image on the negative increases with the exposure time.

No single one of the methods described above gives satisfactory accuracy. Thus, it is very important that the most significant systematic errors be eliminated during processing of the observation results. The commonest processing techniques are the three following: 1) a constant irradiation correction c is obtained for a given set of measurements by measuring an artificial disk with an accurately known angular diameter; 2) corrections are obtained on the basis of theoretical considerations concerning the origin of the errors to be eliminated; and 3) the results of a fairly extensive series are represented by arbitrary equations of the form

$$\delta'' = \frac{\delta_0''}{\Delta} + c, \qquad (3)$$

where δ'' is the measured semidiameter, Δ is the geocentric distance of the body, δ_0'' is the true value of δ for $\Delta = 1$, and c is a constant irradiation correlation. Solving the system of equations by the method of least squares gives the unknown quantities δ_0'' and c.

TABLE 1

Planet	Symbol	Average distance from sun: AU	Average distance from sun: ×10⁶ km	Sidereal revolution (sidereal years)	Mean diameter: km	Mean diameter: relative (earth = 1)	Oblateness, ε	Axial rotation: period	Axial rotation: inclination of equator to orbit	Number of satellites
Mercury	☿	0.3871	57.85	0.2408	5 000	0.39	0	88^d	7° ?	0
Venus	♀	0.7233	108.1	0.6152	12 400	0.97	0	?	14° ?	0
Earth	♁	1.0000	149.4	1.0000	12 742	1.00	1/297	23^h 56^m	23°27′	1
Mars	♂	1.5237	227.7	1.8808	6 770	0.53	1/192	24 37	25 10	2
Jupiter	♃	5.2028	777.6	11.862	140 000	10.95	1/15	9 50	3 7	12
Saturn	♄	9.5388	1426	29.457	115 000	9.02	1/10	10 14	26 45	10
Uranus	♅	19.1910	2868	84.013	51 000	4.00	1/14	10 49	98	5
Neptune	♆	30.0707	4494	164.783	50 000	3.92	1/40	15	29	2
Pluto	♇	39.46	5894	294.	14 000?	1.1?	?	6 4	?	0

Summaries of diameter measurements for various planets have been published repeatedly. Table 1 gives Russell's data, which, although they are not the most recent, are still no less reliable than more recent results,

while they have the additional advantage that the various characteristics of each planet are completely consistent with one another.

§69. **Phases.** Since the part of the surface of a celestial body not illuminated directly by solar rays is generally invisible against the background of the night sky, therefore the apparent outline of the disk is determined by the projection of the illuminated portion only. The line along the surface of the celestial body which separates the illuminated part from the dark part is called the terminator. The terminator is obviously the locus of tangent points between the surface of the body and a cone of rays with its vertex at the center of the sun.

In this discussion of phases we will consider only the case of a spheroidal planet. This is sufficient for practical purposes, since all three of the celestial bodies for which the entire range of phases can be observed from the earth (namely the moon, Mercury, and Venus) have negligible oblateness. Moreover, in our approximate solution of the problem the sun will be regarded as a point source of light located at an infinite distance from the illuminated body. Consequently, it will be possible to consider a cylindrical beam of rays parallel to the planet-sun direction, a case analogous to that depicted in the bottom drawing of Figure 112, the only difference being that now in the direction H we have the sun rather than the earth (the observer).

It is obvious that, if these assumptions are made, the terminator is a great circle on the surface of the planet, the plane of the circle being perpendicular to the radius vector of the planet. Its projection onto the image plane is then a half ellipse, whose semimajor axis is equal to the radius of the disk and whose semiminor axis varies with the phase.

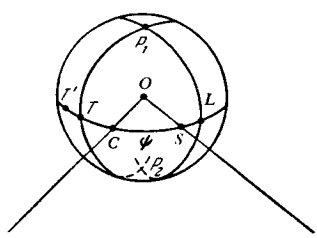


FIGURE 113

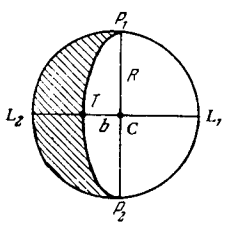


FIGURE 114

The main numerical characteristic of the phase is the phase angle ψ. This is the angle at the planet center O between the direction to the sun and the direction to the observer (Figure 113). The intersection points of these two directions with the surface of the body are: a) the pole of illumination (or the subsolar point) S, for which the sun is at the zenith; and b) the center of the disk C. The great circle on the body surface passing through these points and lying in the plane defined by the planet, the observer, and the sun is called the intensity equator. It is clear that the arc $SC = \psi$, that the limb P_1LP_2 is the polar of the point C, and that the terminator P_1TP_2 is the polar of the point S, so that $\smile CL = \smile ST = 90°$. The points P_1 and P_2 at which the limb meets the terminator both on the body surface and at the apparent disk edge are called the cusps. The disk diameter

P_1P_2 (Figure 114) joining the cusps is called the line of cusps; the diameter L_1L_2 perpendicular to the line of cusps is the projection of the intensity equator. The latter is divided by the terminator P_1TP_2 into the illuminated part L_1T and the dark part TL_2. The phase of a planet is defined as the ratio

$$\Phi = \frac{L_1T}{L_1L_2}. \tag{4}$$

It follows from elementary geometry that the semiminor axis CT of the terminator ellipse is related as follows to the phase angle ψ:

$$b = CT = R\cos\psi, \tag{5}$$

so that for the phase we have

$$\Phi = \frac{R + R\cos\psi}{2R} = \frac{1}{2}(1 + \cos\psi) = \cos^2\frac{\psi}{2}. \tag{6}$$

This formula leads to the succession of phases, or planetary aspects, listed in Table 2.

TABLE 2

ψ	Name of phase or aspect	b/R	Φ
0°	True inferior conjunction (new moon)	-1	0
$0° < \psi < 90°$	Crescent	$-1 < \frac{b}{R} < 0$	$0 < \Phi < \frac{1}{2}$
90°	Dichotomy (first and last quarters)	0	$\frac{1}{2}$
$90° < \psi < 180°$	Gibbous aspect	$0 < \frac{b}{R} < +1$	$\frac{1}{2} < \Phi < 1$
180°	True opposition (full moon)	$+1$	1

Observations of planetary phases involve measurements of the arcs P_1P_2 and L_1T (using a micrometer or some other instrument), followed by the calculation of the apparent phase Φ' from formula (4). A comparison with the true phase Φ calculated from (6) and listed in astronomical almanacs often gives the result $\Phi' > \Phi$, obviously due to the influence of twilight phenomena in the planetary atmosphere; here the light is scattered partly onto the night hemisphere and the apparent terminator is displaced over the arc section $T'T$ of the intensity equator (Figure 113), called the twilight arc. The opposite is also possible, in which case the phase Φ' obtained observationally is less than the true phase; this corresponds to a negative twilight arc.

§70. Planetographic coordinates of surface points. Since all planets rotate about their axes, it is possible to locate the poles of the rotation axis, the equator, meridians, and parallels on the surface of the planet. In addition, the positions of points on the surface can be specified in terms of planetographic coordinates, the latitude β and the longitude λ, these being analogous to geographic latitudes and longitudes on the earth.

On the planets, just as on the earth, positive latitudes are reckoned in the direction of the north pole (that is, the pole nearer to the north pole of the ecliptic), and positive longitudes are reckoned from some arbitrarily chosen initial meridian in the sense opposite to the sense of planetary rotation.

The meridian passing at a given time through the disk center C is called the central meridian. Due to the rotation of the planet, its longitude in general increases monotonically with time, but it must be remembered that this phenomenon is also influenced by the relative motion of the earth and the planet in space. It is easy to show that, since the observer always remains in the plane of the central meridian, the latter is projected onto the image plane as a straight line constituting one of the disk diameters.

In order to obtain the coordinates β and λ from observation, we must introduce a system of auxiliary quantities called the physical coordinates, which specify the apparent position of the coordinate grid on the image plane. The main physical coordinates published in the astronomical ephemerides are the following.

The position angle P of the projection of the rotation axis and the corresponding projection of the central meridian; this coordinate is reckoned counterclockwise from the northern direction of the declination circle passing through the disk center C, that is, eastward.

The planetographic latitude D (sometimes B) of the disk center C; this corresponds to the planetocentric declination of the earth relative to the planetary equator.

The longitude L of the central meridian relative to the zero meridian.

Finally, the physical coordinates also include the planetographic latitude and longitude of the subsolar point S.

In order to determine the planetographic coordinates, a micrometer or some other instrument is used to measure the rectangular coordinates x and y of a surface detail in the image plane. Here the disk center C is taken as the origin and the axis OY runs along the northward direction of the declination circle through the point C. Then, using the formulas

$$\varrho'' = \sqrt{x^2+y^2} \quad \text{and} \quad \operatorname{tg} U = \frac{y}{x},$$

we convert to polar coordinates and calculate the planetocentric angle γ between the directions to the observer and to the detail. This angle is found from the expression

$$\sin(\gamma+\varrho) = \frac{R}{\Delta}\sin\varrho,$$

obtained from the plane triangle through the surface detail, the center of the planet, and the observer. Here Δ denotes the side of the triangle joining the observer and the center of the planet, and R is the radius of the planetary sphere.

In order to convert from the coordinates U, γ to the required planetographic coordinates β, λ, let us consider a spherical triangle on the planetocentric sphere; the vertexes of the triangle are the north pole P of the planet, the disk center C, and the projection M of the detail (Figure 115). Since the sides of this trangle have the lengths $PC = 90° - D$, $PM = 90° - \beta$, and $CM = \gamma$, while the angles at points P and C are equal to $\lambda - L$ and $P + U$,

respectively, it is easy to show that

$$\left.\begin{aligned} \sin\beta &= \sin D\cos\gamma + \cos D\sin\gamma\cos(P+U),\\ \cos\beta\cos(\lambda-L) &= \cos D\cos\gamma - \sin D\sin\gamma\cos(P+U),\\ \cos\beta\sin(\lambda-L) &= \sin\gamma\sin(P+U), \end{aligned}\right\} \qquad (7)$$

which completes the solution of our problem.

The accuracy in determining the planetographic coordinates is usually low, so that it is preferable to replace the trigonometric calculation by various simplified graphical and computational techniques, these being available in great variety.

The results of determining the coordinates of many planetary surface details have been arranged into catalogs, or even into maps or globes of the celestial-body surface. It should be noted that the planetographic coordinates pertaining to different planets are generally given a designation incorporating the Greek name of the particular celestial body. For example, the latitudes and longitudes of lunar features are called s e l e n o g r a p h i c (from the Greek selene, meaning moon), while the coordinates of Martian features are called a r e o g r a p h i c (from Ares, the name for Mars).

§71. **The rotation elements and their observational determination.** Observations of the apparent motions of spots and surface details over the disk of a planet make it possible to determine the elements specifying the rotation of the planet about its axis. These rotation elements are: 1) the inclination i of the equatorial plane of the planet, with respect to its orbital plane; 2) the longitude Ω of the ascending node of the equator, which is the point of vernal equinox for the particular planet; and 3) the sidereal rotation period T of the planet, or else the angular rotation velocity $\xi = 360°/T$ expressed in degrees per day. In some cases the elements i and Ω are related to the plane of the ecliptic and not to the orbital plane of the planet; in other cases they are replaced by the ecliptic or equatorial coordinates of the north pole of the planet on the celestial sphere.

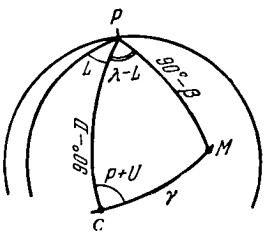


FIGURE 115

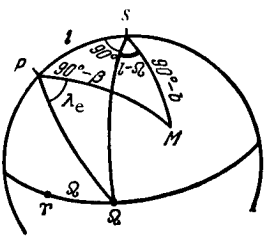


FIGURE 116

A knowledge of these rotation elements completely solves the problem of the succession of days and seasons on a given planet. The sidereal rotation period (neglecting precession) is in effect equal to the length of the sidereal day for the given planet; the mean solar day is longer by an amount equal to the mean diurnal orbital motion of the planet, expressed in units of time.

The inclination i of the equator to the plane of the orbit determines the nature of the seasonal change. The amplitude of the seasonal change is astronomically defined as the difference between the midday heights of the sun at some point on the planet surface during the summer and winter

solstices; this difference for the temperate zone is equal to $2i$. The element i also determines the positions of climatic zones on the planet. Since the tropics are parallels for which the midday zenith distance of the sun at the corresponding solstices is zero, the latitude of these parallels is i; analogously, the latitude of the polar circles is $90° - i$.

The beginning of each season is determined by the planetocentric longitude of the sun relative to the plane of the planetary orbit. If these longitudes are reckoned from the corresponding point of vernal equinox, that is, from the ascending node of the planetary equator along the orbit, then the vernal equinox corresponds to 0° longitude of the sun, the summer solstice corresponds to 90°, the autumnal equinox to 180°, and the winter solstice to 270°.

Precessions of the planetary axes are determined theoretically, since observational determinations of Ω are not accurate enough. Given the masses and the orbital elements of the satellites, and also the oblateness of the planet, we can find the total precession due to the perturbing influence of the satellites and the sun, on the basis of the general theory of precession developed in celestial mechanics.

If the rotation elements of a planet are not known, then it is possible to determine for some surface detail the planetocentric coordinates relative to the plane of the ecliptic. Let the latitude of the detail in this system be b and let the longitude be l, the latter being measured from the point of earth's vernal equinox. Next, let us consider the spherical triangle SPM (Figure 116), whose vertexes are the north pole S of the ecliptic, the north pole P of the rotation axis of the planet, and the projection M of the detail. In this triangle, $MP = 90° - \beta$, $MS = 90° - b$, $PS = i$ (this is the inclination of the planetary equator to the ecliptic), $\angle MSP = 90° + (l - \Omega)$, Ω being the longitude of the ascending node of the planetary equator on the ecliptic, and $\angle MPS = 90° - \lambda_e$, λ_e being the longitude of the detail relative to the planetary equator and reckoned from the ascending node. Consequently, we obtain

$$\sin\beta = \sin b\cos i - \cos b\sin i\sin(l-\Omega), \tag{8}$$

$$\left.\begin{aligned}\cos\beta\sin\lambda_e &= \sin b\sin i + \cos b\cos i\sin(l-\Omega),\\ \cos\beta\cos\lambda_e &= \cos b\cos(l-\Omega)\end{aligned}\right\} \tag{9}$$

If we next make three successive observations of the detail, for three different times t_1, t_2, and t_3, we will obtain a system of three equations of the form (8). These can then be solved for the three unknowns, namely the rotation elements i and Ω and the latitude β of the planetary detail. The rotation velocity ξ of the detail can also be found as the ratio

$$\xi = \frac{\lambda_{3e} - \lambda_{1e}}{t_3 - t_1}. \tag{10}$$

At present, however, this classical means of determining the rotation elements is seldom used. The widespread use of photographic observations has made the method proposed by the American astronomer Trumpler very popular. According to this method, a series of successive photographs of the planet is used to find the rectangular coordinates of a surface detail in the image plane. Then, applying the method of least squares or by some graphical means, we establish the apparent trajectory of the detail, which is a semiellipse. Since this semiellipse is the projection of a planetographic parallel, therefore its minor axis will be oriented along the projection of

the rotation axis, its eccentricity will give the inclination of the equator with respect to the line of sight, and the semimajor axis will indicate the latitude of the detail. The advantage of this method is that it minimizes the influence of systematic errors arising due to photographic irradiation when the cross hairs of the measuring instrument are adjusted to the edge of the image.

The rotation elements are now known for all the planets, except for Venus and Pluto. Thus, future research in this respect will be dedicated to improving the accuracy of the figures obtained. By differentiating equation (8), we obtain

$$\cos\beta\, d\beta = -\sin b \sin i\, di - \cos b \sin(l - \Omega) \cos i\, di + \sin i \cos b \cos(l - \Omega)\, d\Omega.$$

Then, by means of (9), we can simplify this to

$$d\beta = \cos\lambda_e \sin i\, d\Omega - \sin\lambda_e\, di. \tag{11}$$

If finite increments are substituted for the differentials, while the latitude of the detail corresponding to the initial set of elements is denoted by β, and the latitude corresponding to the corrected elements is denoted by β_0, then we can define

$$\beta - \beta_0 = \Delta\beta, \quad \Delta\Omega = \sin i = X, \quad \Delta i = -Y.$$

Consequently, equation (11) can be rewritten as

$$\beta = \beta_0 + \Delta\beta = \beta_0 + X\cos\lambda_e + Y\sin\lambda_e. \tag{12}$$

If the values taken for the elements are considerably different from the actual figures, then the corresponding β will vary sinusoidally with λ_e. Once we have established this sinusoidal variation, it is easy to find the corrections $\Delta\Omega$ and Δi.

In order to find the correction $\Delta\xi$ to the initial angular velocity of rotation ξ, it is necessary to make several successive measurements of the longitude λ of some detail relative to the zero meridian, which rotates with a velocity ξ. A regular variation with time of the measured λ indicates the unsuitability of the adopted value of ξ. By expressing the observational results in the form of equations such as

$$\lambda = \lambda_0 + \Delta\xi\,(t - t_0), \tag{13}$$

where λ_0 is the longitude at the initial time t_0, and by solving this system using the method of least squares, it is possible to find the rotation-velocity correction $\Delta\xi$.

§72. **Spectroscopic studies of rotation.** An alternative means of studying planetary rotations is based on an application of the Doppler-Fizeau effect. For a celestial body which only reflects the incident solar radiation, the shift of the spectral lines is determined both by the velocity $v_E = \frac{dR_E}{dt}$ of the body relative to the earth and by its velocity $v_S = \frac{dR_S}{dt}$ relative to the sun (here R_E and R_S are the geocentric and heliocentric distances of the body). The

variation $\Delta\lambda$ (of the wavelength λ of the spectral line) corresponding to these velocities is given by the general formula

$$\Delta\lambda = \frac{\lambda}{c}(v_E + v_S), \tag{14}$$

where c is the velocity of light, and the velocities v_E and v_S are taken as positive for receding motion and negative for approaching motion.

The shift $\Delta\lambda$ is primarily due to the orbital motions of the earth and the planet; obviously, this part of the shift will be the same for all points on the planet disk. However, the rotation of the planet about its axis also has some effect, and this of course varies for different points on the disk.

It is easy to see that the effect of rotation will be a maximum for the points at the edge of the disk which lie along the equator. Let us assume a phase angle of zero, so that the entire disk is illuminated. Due to the rotation, points lying on the planetary equator move with a velocity v_0, defined as

$$v_0 = \frac{2\pi R}{T}, \tag{15}$$

where R is the radius of the equator, and T is the rotation period. Since the planets as a rule have direct rotation, the point of intersection of the equator with the eastern limb approaches the observer and the sun with a velocity v_0, while the point on the western limb recedes with the same velocity; thus, a corresponding shift of the spectral lines is produced. This is the case, however, only when the earth and the sun lie in the equatorial plane of the planet. Let us consider a general case of relative orientation of the observer, the planet, and the sun.

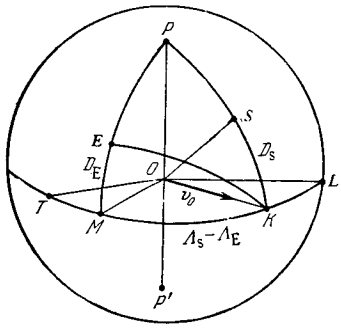


FIGURE 117

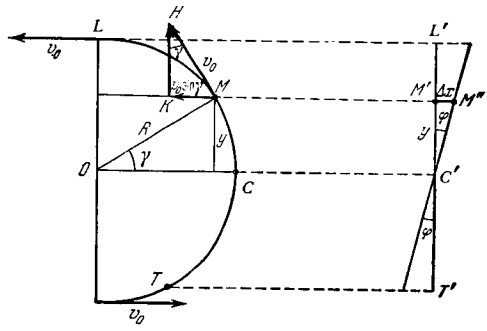


FIGURE 118

In Figure 117 O is the center of the planetocentric sphere, PP' is the rotation axis of the planet, $TMKL$ is the planetary equator, S and E are the projections of the sun and the earth, and T and L are the intersection points of the equator with the terminator and the limb. If we construct the projections PEM and PSK of the planetographic meridians, then we have $\smile EM = D_E$, $SK = D_S$, and $\smile KM = \Lambda_S - \Lambda_E$, where Λ_S and Λ_E are the planetographic longitudes of the sun and the earth. Next, let us find the shift corresponding to the point T. The equatorial velocity v_0 at this point, which is directed

perpendicular to the radius OT of the sphere, is parallel to the radius OK. By projecting this velocity onto the directions OS and OE, we can find the velocity components v_{TS} and v_{TE}:

$$v_{TS} = -v_0 \cos D_S \quad \text{and} \quad v_{TE} = -v_0 \cos D_E \cos(\Lambda_S - \Lambda_E).$$

Consequently, according to (14), the line shift $\Delta\lambda_T$ is

$$\Delta\lambda_T = -v_0 \frac{\lambda}{c} [\cos D_S + \cos D_E \cos(\Lambda_S - \Lambda_E)]. \tag{16}$$

Analogously, the velocity of rotation at point L on the limb is parallel to the radius MO perpendicular to OL. By projecting the radius MO onto the directions OE and OS, we can find the velocity components v_{LE} and v_{LS} for the point L:

$$v_{LE} = +v_0 \cos D_E \quad \text{and} \quad v_{LS} = +v_0 \cos D_E \cos(\Lambda_S - \Lambda_E),$$

which gives a spectral-line shift $\Delta\lambda_L$ of

$$\Delta\lambda_L = v_0 \frac{\lambda}{c} [\cos D_E + \cos D_S \cos(\Lambda_S - \Lambda_E)]. \tag{17}$$

In practice, the observations are generally set up so that the shift $\Delta\lambda$ of lines at point L relative to point T are obtained. This relative shift is

$$\Delta\lambda = \Delta\lambda_L - \Delta\lambda_T = v_0 \frac{\lambda}{c} (\cos D_E + \cos D_S)[1 + \cos(\Lambda_S - \Lambda_E)]. \tag{18}$$

The best measurements are those made near opposition. In the ideal case, when $D_E = D_S = 0$ and $\Lambda_S - \Lambda_E = 0$, we obtain

$$\Delta\lambda = 4v_0 \frac{\lambda}{c}. \tag{19}$$

The presence of the multiplying factor 4 facilitates the measurements, since even for small velocities v_0 it ensures an appreciable shift $\Delta\lambda$.

Given v_0, we can find the rotation period of the planet from the formula

$$T = \frac{2\pi R}{v_0}, \tag{20}$$

where R is the radius of the equator.

The observation technique is as follows: first, the image of the planetary disk is projected onto the slit of a high-dispersion spectrograph with the aid of a long-focus objective or mirror. Here, the spectrograph is set so that the slit is parallel to the projection of the meridian section TL. On the spectrogram the effect of the planetary rotation shows up as a movement of the ends of the lines in opposite directions, and as a result the lines appear tilted somewhat relative to their normal direction (Figure 118, left); this movement of the line ends takes place because these ends correspond to the spectra of points T and L of the disk. The following three methods can be used to obtain the velocity v_0 from such a spectrogram.

1. Absolute measurements of the line-of-sight velocities v_T and v_L for points T and L can be made by the methods described in Volume I, and then the equatorial velocity can be obtained from (16) and (17). The shortcomings

of this method are: it is laborious; high systematic errors are introduced because of the blurred edges of the spectrum on the spectrogram, due to image vibration and photographic irradiation; the settings of the relevant points on the disk are inaccurate (actually, we measure not points T and L but rather parts of the disk which are displaced more toward the center); instrument effects enter in.

2. A relative comparison of line positions at the opposite ends of a single spectrum or for two different spectra is made. When carried out with a spectrocomparator, such a comparison gives the linear relative displacement Δx of the line ends, and also, in terms of the dispersion $D = \frac{d\lambda}{dx}$, the corresponding wavelength difference $\Delta\lambda$ expressed by (18). The errors introduced are the same as in the preceding method.

3. Measurements of the line tilt are made. In order to explain this method, let us consider Figure 118, where TCL is the half of the equator facing the observer, M is some point on it, and MH is the vector of velocity v_0. For an observer situated in the direction OC, point M is projected onto the image plane $T'L'$ as point M', located a distance $y = R\sin\gamma$ from C'. The quantity R is the radius of the equator, and γ is the planetocentric angular distance of point M from the disk center C. The component of the velocity v_0 along the line of sight MM' is the vector MK, which has an amplitude of $v_{0E} = v_0 \sin\gamma$. Thus, $v_{0E} = v_0 \frac{y}{R}$, and the linear shift of the spectral line at point M' of the spectrogram (corresponding to point M of the disk) is

$$M''M' = \Delta x = D\Delta\lambda = v_0 \frac{D}{R}\frac{\lambda}{c} y.$$

It is clear that the shift Δx is proportional to the distance y from the central line of the spectrogram, corresponding to the point C' (Figure 118, right). The rotational Doppler effect thus turns the line through some angle φ relative to its normal direction $L'C'$, without distorting the line.

From the figure we see that

$$\operatorname{tg}\varphi = \frac{\Delta x}{y} = v_0 \frac{D}{R}\frac{\lambda}{c},$$

so that the velocity v_0 is

$$v_0 = c\,\frac{R\varphi^\circ \sin 1^\circ}{\lambda D}. \qquad (21)$$

The angle φ is determined using an instrument which measures position angles. The cross hairs of the measuring microscope are first set parallel to the unshifted lines of a comparison spectrum (the direction $M''C'$), and then to the tilted lines of the planetary spectrum. The difference in position-angle readings gives the angle φ. This method is not very accurate as far as random errors are concerned, but on the other hand it is more or less free of the systematic errors which characterize the other methods.

The Doppler effect was first utilized as a means for studying planetary rotations by Belopol'skii at Pulkovo Observatory. He used spectrograms taken with the 30-inch refractor to determine successfully the rotation periods of Jupiter and Saturn, and also the periods of various zones in Saturn's rings.

§73. The disk of a considerably flattened planet. Let us consider a planet which is a spheroid, with an oblateness ε and a corresponding

meridional eccentricity e. If the observer is situated in the equatorial plane of the planet, that is, if $D = 0$ (see Figure 119), then the planetary disk will appear as an ellipse with semiaxes a and b and eccentricity e. As D increases, the eccentricity decreases, until for $D = 90°$ we have $e = 0$, since in this case the limb coincides with the equator (the situation actually observed for Uranus).

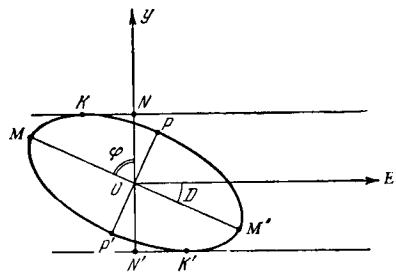


FIGURE 119

In order to obtain a relation between e and D, let us consider Figure 119. Here the ellipse $PMP'M'$ represents the central meridian of the planet, $OP = b$ and $OM = a$ are the semiaxes of the meridian, OE is the direction of the terrestrial observer, and $\angle EOM = D$. Let us also draw the tangents NK and $N'K'$ to the ellipse. Obviously, the distance NN' between the two tangents gives the polar diameter of the disk on the image plane.

Next we take a system of rectangular coordinates with the origin at the center of the ellipse O and with the axes OX and OY pointing along the semimajor and semiminor axes of the ellipse. The coordinates of the tangent point K are x_1 and y_1. Since we know the equation for the line tangent to an ellipse at a given point, we have

$$x\frac{x_1}{a^2} + y\frac{y_1}{b^2} - 1 = 0. \tag{22}$$

On the other hand, since $ON \perp NK$, we can set $NO = b_1$ and $\angle YON = \varphi$ to obtain the equation for this line in "normal" form:

$$x\cos\varphi + y\sin\varphi - b_1 = 0$$

or

$$x\frac{\cos\varphi}{b_1} + y\frac{\sin\varphi}{b_1} - 1 = 0.$$

By comparing the coefficients of x and y in this equation and in (22), we obtain the equalities

$$\frac{x_1}{a^2} = \frac{\cos\varphi}{b_1} \quad \text{and} \quad \frac{y_1}{b^2} = \frac{\sin\varphi}{b_1}$$

or

$$\frac{x_1}{a} = \frac{a}{b_1}\cos\varphi \quad \text{and} \quad \frac{y_1}{b} = \frac{b}{b_1}\sin\varphi.$$

By squaring these expressions and adding them up, we obtain

$$\frac{x_1^2}{a^2} + \frac{y_1^2}{b^2} = \frac{a^2}{b_1^2}\cos^2\varphi + \frac{b^2}{b_1^2}\sin^2\varphi = 1.$$

Then, solving this equation for b_1 and taking into account that $b^2 = a^2(1-e^2)$, we find that

$$b_1^2 = a^2\cos^2\varphi + b^2\sin^2\varphi = a^2(1 - e^2\sin^2\varphi).$$

Applying the equalities

$$e^2 = 2\varepsilon - \varepsilon^2 \quad \text{and} \quad \varphi = 90° - D,$$

we finally obtain for the polar semidiameter b_1 of the disk

$$b_1 = a\sqrt{1 - (2\varepsilon - \varepsilon^2)\cos^2 D}. \tag{23}$$

The apparent oblateness ε_1 of the disk is therefore

$$\varepsilon_1 = \frac{a - b_1}{a} = 1 - \sqrt{1 - (2\varepsilon - \varepsilon^2)\cos^2 D}. \tag{24}$$

Since ε is small, whereas the accuracy with which a and b can be determined in practice is still not high, it is advisable to drop the terms of second and higher orders in ε. Consequently, we obtain

$$\varepsilon_1 = 1 - (1 - 2\varepsilon\cos^2 D)^{1/2} = \varepsilon\cos^2 D.$$

According to these formulas, the expression giving the true oblateness ε of a spheroidal planet, in terms of the measured apparent oblateness ε_1 of its disk, has the form

$$\varepsilon = \varepsilon_1 \sec^2 D. \tag{25}$$

§74. **Planetary characteristics related to the mass.** The mass is the principal characteristic of any celestial body. Problems of mass determination are discussed in texts on celestial mechanics, and here it will be enough to mention that two methods exist for finding the mass. The first method is based on the application of Kepler's third law, as generalized by Newton to describe the motions of satellites revolving about a given planet. Its application requires that a detailed theory of the motion of the satellite be available, since only such a theory can provide the basic parameters needed for the calculation, namely the semimajor axis a of the orbital ellipse and the revolution period T. The second method is based on an analysis of the perturbations which the given body exerts on the motions of other bodies. This method is mainly applied to determine the masses of planets which have no satellites, and also to the large satellites of the Jovian planets. In addition, it is often used in conjunction with the first method for mass determinations of planets with satellites. The masses of the smaller members of the solar system, such as the asteroids and small satellites, cannot be determined, so that we must be satisfied with hypothetical estimates based on various indirect considerations.

Given the mass, we can easily find the mean density $\varkappa$ of the body by dividing the mass M by the volume V:

$$\bar{\varkappa} = \frac{M}{V}. \tag{26}$$

For a spherical body, the volume is calculated using the common formula $V = \frac{4}{3}\pi R^3$. For an appreciably flattened body, the volume can be calculated

using the same formula, but in this case R represents the average radius $\bar{R} = \frac{1}{3}(2a + b) = a(1 - \frac{1}{3}\varepsilon)$, where a and b are the semimajor and semiminor axes of the spheroid, and ε is the oblateness; alternatively, the usual ellipsoid formula can be used, namely $V = \frac{4}{3}\pi a^2 b$.

Furthermore, knowing the mass, we can calculate the mean acceleration of gravity at the surface of the planet:

$$g = G\frac{M}{R^2}, \tag{27}$$

where G is the gravitational constant, equal to $6.67 \cdot 10^{-8}$ CGS units. The distribution of the force of gravity over the surface of the planet can be expressed by the familiar Clairaut formula:

$$g = g_0(1 + \eta \sin^2 \beta), \tag{28}$$

where β is the planetographic latitude, g_0 is the acceleration of gravity at the equator (expressed to a first approximation by (27)), and the parameter η has the value

$$\eta = \frac{g_{90^\circ} - g_0}{g_0} = \frac{5}{2}q - \varepsilon.$$

Here q is the ratio between the centrifugal acceleration at the equator and g_0. For our purposes, this ratio can be taken as approximately equal to

$$q = \frac{3}{4}\frac{\omega^2}{G\pi\bar{\varkappa}} = \frac{3\pi}{G\bar{\varkappa}T^2}, \tag{29}$$

where ω is the angular rotation velocity of the planet, and T is the period of its axial rotation. The parameter q is significant in that, in conjunction with the oblateness ε, it makes possible certain conclusions concerning the internal structure of the body. Specifically, Clairaut showed that the ratio ε/q for any real body must lie between $1/2$ and $1\,1/4$. The lower value corresponds to the case of a limit concentration of matter at the body center, when almost all the mass is concentrated at the center, while the upper value corresponds to a constant density of matter throughout the body.

TABLE 3

Planet	$\frac{M}{M_E}$	$\bar{\varkappa}$, g/cm	$\frac{\bar{\varkappa}}{\varkappa_E}$	$\frac{g}{g_E}$	$\frac{\varepsilon}{q}$	v_e, km/sec
Mercury	0.04	3.8	0.70	0.27	—	3.6
Venus	0.81	4.9	0.88	0.85	—	10.2
Earth	1.00	5.5	1.00	1.00	0.97	11.2
Mars	0.11	4.0	0.72	0.38	1.14	5.0
Jupiter	316.9	1.3	0.24	2.64	0.63	60
Saturn	94.9	0.7	0.13	1.17	0.54	36
Uranus	14.7	1.3	0.23	0.92	0.74	21
Neptune	17.2	1.6	0.29	1.12	0.57	23
Pluto	0.99					

Another important physical characteristic of a body is its critical velocity, or escape velocity, v_e. This is the velocity at which motion near

the surface of the planet is parabolic *:

$$v_e = \sqrt{2G\frac{M}{R}} = \sqrt{2g\bar{R}}. \tag{30}$$

The different mass characteristics discussed above are given in Table 3.

Here $M_E, \varkappa_E$ and g_E denote the mass, mean density, and acceleration of gravity for the earth.

§75. The brightness, magnitude, and color of a planet. For planets and satellites, as well as for stars (see Volume I, § 122), the illumination E_p caused by the body at the point of observation is called the brightness. The decimal logarithm of the brightness multiplied by −2.5 gives the stellar magnitude of the body m_p. This quantity is measured by means of a photometric comparison with stars of known magnitude (north-polar sequence, standard areas), the technique used being that of stellar photometry. Difficulties sometimes arise which require the application of special measures. On the one hand, these may be due to the proper motions of planets and satellites, as a result of which their images on the negatives will be elongated, making comparisons with the stars impossible. On the other hand, they may be due to the considerable angular extent of the disks of the major planets.

The brightness and magnitude of a planet vary in time fairly rapidly, mainly due to the variation of its geocentric and heliocentric distances and to its phase variation. Some influence may also be attributed to other factors, such as the inclination of the equatorial plane to the line of sight, the axial rotation (which alters the relative positions of light and dark spots on the disk), the seasonal changes in surface color, etc., but the effects of these are secondary.

Let us consider a spherical planet of radius R. Its brightness E_p is proportional to the projection area of the sphere πR^2 and to the solar illumination E_s at a certain standard distance Δ_0; it is inversely proportional to the squares of the radius vector Δ and the geocentric distance L. Taking Γ as the proportionality coefficient, we may write

$$E_p = \Gamma E_s \frac{\Delta_0^2}{\Delta^2}\frac{\pi R^2}{L^2} f(\psi). \tag{31}$$

Here $f(\psi)$ is the so-called phase function or phase curve, which gives the variation of brightness with the phase angle ψ; it is normalized so that $f(0) = 1$. Now, in terms of stellar magnitudes, we can write (31) as

$$m_p = -2.5 \lg E_p = -2.5 \lg \pi\Gamma E_s \Delta_0^2 R^2 + 5 \lg \Delta L - 2.5 \lg f(\psi). \tag{32}$$

The quantity Δ_0 is generally taken as one astronomical unit, E_S is taken as the solar illumination at the mean earth-sun distance, and Δ and L are also expressed in astronomical units. In this case the first term on the right side of (32),

$$g = -2.5 \lg \pi\Gamma E_s R^2, \tag{33}$$

* The velocity v_e at which a body escapes from the surface of a planet to infinity is known [in Soviet literature] as the second cosmic velocity. The first cosmic velocity $\sqrt{g\bar{R}}$ [more commonly known in the West as the circular velocity] is the velocity at which a body becomes a satellite of the planet and revolves in a circular orbit of radius R.

obviously represents the stellar magnitude of the planet as viewed by an observer at the center of the sun, provided he is a distance of $\Delta = 1$ A.U. from the planet. The quantity g is called the absolute magnitude of the satellite; this is a basic photometric constant for any celestial body. In other cases, albeit only for the superior planets, the stellar magnitude at mean opposition m_0 is given. This is the magnitude which the planet would have for $\psi = 0$, if its distance from the sun were equal to the semimajor axis a of the orbit and its distance from the earth were $L = a - 1$. It is easy to see that m_s and g are related by the expression

$$m_s = g + 5 \lg a (a - 1). \tag{34}$$

The phase curve can be plotted empirically by measuring the magnitude of the planet at various phase angles ψ. Such measurements, of course, can be made with much completeness only for three bodies, Mercury, Venus, and the moon, since the superior planets can be observed only within a limited range of ψ. The results of the measurements are usually represented as a power series;

$$-2.5 \lg f(\psi) = k\psi + l\psi^2 + n\psi^3 + \dots$$

For the superior planets, for which observations give only the initial section of the curve, we retain only the first term. The coefficient k in this term is called the phase coefficient. This coefficient is one of the principal photometric characteristics of celestial bodies.

From the foregoing, we obtain the following formula for m_p:

$$m_p = g + k\psi + l\psi^2 + n\psi^3 + 5 \lg \Delta L. \tag{35}$$

Formulas of this kind are used, in particular, to calculate the values of m_p published in astronomical almanacs. The numerical values of parameters g, m_0, and k, and also the photometric characteristics of planets to be considered below, are given in Table 4.

TABLE 4

Planet	m_0	g	k	δ_1	A_g	A_i	q	A_s	Color index
Mercury I	−2.94	−0.88	0.037	3".45	0.164	0.246	0.42	0.069	
Mercury II	−2.12	−0.06	0.023	3.45	0.077	0.115	0.72	0.055	
Venus	−4.77	−4.06	0.013	8.55	0.492	0.738	1.20	0.59	+0.78
Earth	−4.57*	−3.52*	0.016*	8.79*	0.29*	0.44*			0.45
Mars	−1.85	−1.36	0.015	4.67	0.139	0.208	1.11	0.154	1.38
Jupiter	−2.29	−8.99	0.015	95.23	0.375	0.562	—	—	0.96*
Saturn	+0.89	−8.67	0.015	77.95	0.420	0.63	—	—	1.12
Uranus	+5.74	−6.98	—	36.00	0.42	0.63	—	—	
Neptune	+7.15	−7.06	—	34.5	0.49	0.74	—	—	
Pluto	+14.71*	−1.19*	—	—	0.03*?	0.04*?	—	—	0.67*

* All the values except those marked with an asterisk * are from Russell

By measuring the magnitude of a planet in two spectral regions, for instance in the photographic and photovisual regions, we can obtain the color index C. Since planets and satellites shine by reflected sunlight, this characteristic is not as significant for them as for stars. A more useful

parameter is the so-called color excess or yellowness index, defined as

$$D = C - C_{\odot}, \tag{36}$$

where C is the color index of the planet, and $C_{\odot}$ is the color index of the sun. Obviously, when $D > 0$, the body surface reflects predominantly the long-wave part of the spectrum; for $D < 0$, on the other hand, the short-wave radiation is scattered strongly; finally, for $D = 0$, we have the case of neutral reflection. In practice, the determination of D is complicated by the fact that it is difficult to obtain the solar color index $C_{\odot}$ in the same color system. Methods have therefore been proposed which enable a direct comparison between the color of a planet and the color of sunlight reflected by a white screen. On the other hand, it is permissible to replace the sunlight by the light of stars in the same spectral class as the sun (dG2); this greatly simplifies the process of colorimetric observation, since the determination of D is reduced to a comparison of the colors of the planet and the star when the two are at equal heights above the horizon.

Results more comprehensive than those of colorimetry are provided by the integrated-light spectrophotometry of planets. Here, the standard may be either sunlight or the light from stars with the same spectral composition. The final result of the spectrophotometric observations, which are made either with a slit spectrograph or an objective-prism camera, is a curve giving the ratio of the planetary-light intensity $j(\lambda)$ to the corresponding sunlight intensity $j_{\odot}(\lambda)$, as a function of the wavelength λ. This ratio is generally specified in relative units, the ratio of intensities for some arbitrary λ being taken as unity.

Another technique for studying the integrated light of a planet or satellite is based on polarization measurements. Reflected light is as a rule partially polarized, that is, it is a mixture of natural unpolarized light of intensity E_n and light polarized in a certain plane, whose intensity we will denote as E_p. The problem in this case is to measure the degree of polarization p:

$$p = \frac{E_p}{E_n + E_p}, \tag{37}$$

and to find the position of the plane of polarization. The technique for making polarization observations was developed most thoroughly by the French astronomer Lyot, who designed a very sensitive visual polarimeter and used it to carry out extensive sets of observations of the light from the moon and the planets. Photographic photometry is less useful for polarization observations, since its low accuracy prevents the detection of the slight polarization which is typical of the light from a planet; however, photoelectric methods can be applied with considerable success here.

§76. Reflectivity. The final aim of photometric observations of planets and satellites is to determine certain numerical characteristics of the reflectivity of their surfaces. These characteristics enable subsequent comparisons with the various materials, rocks, and forms of the earth's landscape.

With respect to incident light, the surfaces of planets and satellites are mat or dispersive. This means that an incident beam of solar radiation is scattered by the surface in all directions. Correspondingly, a plane element of the surface illuminated from outside will have a certain brightness B in

every direction, and this brightness will be proportional to the illumination E of the element:

$$B = \frac{r}{\pi} E. \tag{38}$$

The coefficient

$$r = \frac{\pi B}{E} \tag{39}$$

entering into this formula is called the brightness coefficient or the luminosity; it constitutes the main reflectivity characteristic for a mat surface.

For some scattering substances the brightness is the same in all directions. This means that the coefficient r is independent of the reflection angle ε between the reflected light and the normal to the surface. A case can be imagined in which the value of r also remains the same for all incidence angles i of the illuminating beam. A surface with constant r, namely whose luminosity is the same for all possible combinations of angles of incidence and reflection, is called ideally mat or orthotropic.

No ideally mat materials exist in nature, but the concept of an orthotropic surface is of considerable theoretical significance for photometric studies of reflecting bodies. Many applied photometric problems make it necessary to provide a screen which reflects orthotropically within some given margin of error. Such screens may have dull or mat surfaces of frosted glass, marble, chalk, or gypsum, or better still, floury surfaces made specially by pressing on some powdered substance such as magnesium oxide (MgO) or baryta ($BaSO_4$). However, for angles of incidence close to 90°, the departure from orthotropic reflection will be appreciable even for such screens.

Let us consider an element of an orthotropic surface s illuminated by a uniform parallel beam which causes an illumination E_0 at a surface perpendicular to it. If the angle of incidence at the surface element is i, then the illumination of the surface will be $E_0 \cos i$, while the brightness is given by formula (38) as $\frac{r}{\pi} E_0 \cos i$. Next, by applying the familiar relation between the luminous intensity and the brightness (see Volume I, §117, formula (19)), we obtain the following expression for the luminous intensity I of the element in a direction lying at an angle ε to the normal:

$$I = \frac{r}{\pi} s E_0 \cos i \cos \varepsilon. \tag{40}$$

Thus the luminous intensity of a scattering element is proportional to the cosines of the angles of incidence and reflection. This proposition is known as "the law of cosines" or "Lambert's law"; it obviously applies only for constant r. Therefore, we can also define an orthotropic surface as a surface which strictly obeys Lambert's law of cosines.

For real substances the coefficient r varies with the direction, sometimes very considerably. Let us now introduce a system of spherical coordinates with the pole in the direction of the normal to the element. A particular direction will be specified by its polar distance i or ε and its azimuth ϑ, reckoned from the plane containing the direction of the illuminating beam. The brightness coefficient can thus be written as a function of three angles:

$$r = r_0 f(i, \varepsilon, \vartheta), \tag{41}$$

where r_0 is a constant. Since this function as a rule can be specified only by numerical means, it is best to plot it graphically as a vector diagram, in which segments with lengths proportional to the value of r for a given i are laid off from the origin in different directions. A surface constructed through the ends of these segments is called the reflection indicatrix. Reflection indicatrices may have a great variety of shapes. The two following, however, are the most frequent: 1) an indicatrix elongated in the direction of the illuminating beam toward the source of light; such an indicatrix is characteristic of highly irregular, pitted formations like ploughed soil, vegetation cover, heaped fragments, etc.; 2) an indicatrix elongated in the direction of the specularly reflected ray, which indicates the presence of specular reflection, giving characteristic reflexes and highlights (glossy substances). The indicatrix of an orthotropic surface will obviously be hemispherical. The indicatrices of various materials have been studied in numerous publications by Orlova.

The simplest way to determine the brightness coefficient in practice is to compare the brightness B of the given surface with the brightness B_s of a standard screen under entirely identical illumination conditions. From formula (38) it follows that, if the brightness coefficient r_s of the screen is known, then the value of r for the substance in question can be found from the equation

$$r = r_s \frac{B}{B_s}. \tag{42}$$

When very high accuracy is required, the absolute standard for these measurements is a white screen (for example, one of gypsum), which has been covered with magnesium oxide from the smoke of a flame of chemically pure metallic magnesium. For such a screen, $r_s = 1$ to within 2 or 3%.

The quantity r measured in a monochromatic luminous flux of wavelength λ is designated as $r(\lambda)$ and is called the spectral brightness coefficient. Its variation over the spectrum determines the color of the substance. If $r(\lambda)$ is constant within the relevant spectral range, then the substance is said to be neutrally dispersive or gray.

Spectral brightness coefficients are usually determined using formula (42), the brightnesses in monochromatic light being compared spectrophotometrically. A great many measurements of $r(\lambda)$ for various substances and terrestrial landscapes have been carried out (and the results published) by Krinov.

§77. Albedo. The collective term albedo, which is derived from the Latin word "albus" (meaning white) and which consequently denotes "whiteness," is applied in astronomy to a group of quantities characterizing the overall fraction of light reflected by a given surface in all directions.

The simplest example is the albedo for a homogeneous element of a plane surface. If Φ_0 is the luminous flux incident on the element and Φ is the flux scattered back into space, then the ratio

$$A = \frac{\Phi}{\Phi_0} \tag{43}$$

is called the plane or true albedo. Sometimes it is also called Lambert's albedo, while in illumination engineering it is referred to as the diffuse reflection coefficient, ϱ.

The plane albedo of a given specimen in general varies with the illumination conditions and, in particular, with the angle of incidence i of the illuminating flux at the surface. It also differs for different spectral regions.

Since the incident flux is proportional to the illumination ($\Phi_0 = sE$) and the scattered flux is proportional to the luminosity ($\Phi = sR$), the albedo A will obviously be equal to the ratio R/E. Therefore, by applying the formulas given in Volume I, §116, it is easy to find a relation between the albedo and the brightness coefficient. In order to do this, in formula (17) of §116 the quantity $r_0 E_0 f(i, \varepsilon, \vartheta) \cos i/\pi$ must be substituted for $B_0 f(\varphi, \vartheta)$ and the angle ε must be substituted for φ, to give the value of R. Since the illumination E equals $E_0 \cos i$, it is easy to show that

$$A = \frac{R}{E} = \frac{r_0}{\pi} \int_{\vartheta=0}^{2\pi} \int_{\varepsilon=0}^{\frac{\pi}{2}} f(i, \varepsilon, \vartheta) \sin \varepsilon \cos \varepsilon \, d\varepsilon \, d\vartheta. \tag{44}$$

For an orthotropic surface $f(i, \varepsilon, \vartheta) = 1$, and therefore

$$A = r_0. \tag{45}$$

In some photometric problems the concept of an absolutely white surface is used, this being an orthotropic surface whose albedo is unity. As (45) shows, the brightness coefficient of such a surface is also unity, and this defines the system of brightness coefficients described above. The term "brightness coefficient" should thus be interpreted as the ratio of the brightness B of a mat surface in a given direction to the brightness B_0 of an absolutely white surface under precisely the same illumination conditions:

$$r = \frac{B}{B_0}. \tag{46}$$

As we noted in the previous section, in practice an absolutely white surface is simulated using a screen coated with magnesium oxide, whose albedo in the visible spectrum is from 0.97 to 0.99.

The determination of the albedo A is fairly complicated even for terrestrial specimens, since it involves measuring the luminous flux Φ scattered over a whole hemisphere. Such measurements are made using so-called integrating photometers and albedometers, a good example of the latter being the albedometer developed by Fesenkov. A method of albedo measurement for the natural cover of the earth's surface under field conditions was developed by the Soviet actinometrist Kalitin. Such methods, however, cannot be applied to the surface elements of celestial bodies, and so the plane albedos of these bodies are in general still unknown.

A special reflectivity characteristic of planets and satellites is the so-called spherical albedo A_s (sometimes called the Bond albedo). The spherical albedo is defined as the ratio of the luminous flux Φ_p dispersed by the whole illuminated hemisphere of the planet in all directions to the luminous flux Φ_0 received from the sun by this hemisphere:

$$A_s = \frac{\Phi_p}{\Phi_0}. \tag{47}$$

Let us consider a spherical planet at a distance Δ from the sun. The point O is the center of the planet (Figure 120), and OS is the direction of

the sun. We next draw a sphere KP of arbitrary radius L centered at O, and we consider on it a system of spherical coordinates with the pole at the point P where the sphere surface meets the radius vector OS. Here, a polar distance MP will be defined by the phase angle ψ.

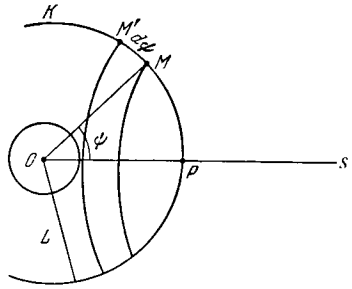


FIGURE 120

Let us now construct on the surface of the sphere an elementary strip MM', bounded by small circles with polar distances ψ and $\psi + d\psi$. The area ds of this strip is $2\pi L^2 \sin\psi \, d\psi$, and the illumination E_p of the surface of the strip, as a result of scattered planetary light, is given by equation (31). Since the luminous flux Φ is equal to the illumination multipled by the area, therefore the element of flux $d\Phi_p$ incident on the strip area will be

$$d\Phi_p = E_p \, ds = 2\pi^2 \Gamma E_s \frac{\Delta_0^2 R^2}{\Delta^2} f(\psi) \sin\psi \, d\psi.$$

The entire flux Φ reflected by the illuminated hemisphere of the planet in all directions can now be obtained by integrating the preceding expression over the entire sphere, which gives

$$\Phi_p = 2\pi^2 \Gamma E_s \frac{\Delta_0^2 R^2}{\Delta^2} \int_0^\pi f(\psi) \sin\psi \, d\psi.$$

On the other hand, the flux Φ_0 incident on the illuminated hemisphere is equal to the solar illumination times the cross-sectional area of the planetary sphere:

$$\Phi_0 = \pi R^2 \frac{\Delta_0^2}{\Delta^2} E_s.$$

After inserting these expression for Φ_p and Φ_0 into (47), we obtain

$$A_s = 2\pi\Gamma \int_0^\pi f(\psi) \sin\psi \, d\psi. \qquad (48)$$

The quantity $\pi\Gamma$ is generally called the geometric albedo; we denote it here as A_g (another symbol for it is p). Twice the value of the integral in (48) is called the phase integral and denoted as q. In terms of this notation, (48) simplifies to

$$A_s = qA_g. \qquad (49)$$

The spherical albedo is thus equal to the geometric albedo multiplied by the phase integral. The latter can be calculated exactly only if the phase curve $f(\psi)$ is known, as is the case for the moon, Mercury, and Venus. Otherwise, approximate methods or arbitrary assumptions are necessary. For Mars, a considerable part of the phase curve is known, so that the remaining part can be obtained by extrapolation. In other cases, empirical relations between the phase coefficient and the phase integral q are used. For the farthest planets, the phase coefficient cannot be determined with much accuracy either. In this case, the spherical albedo is computed assuming an orthotropic planetary surface, that is, formula (40) is used. This gives a phase curve of the form

$$f(\psi)=\frac{1}{\pi}[\sin\psi+(\pi-\psi)\cos\psi], \tag{50}$$

which when substituted into (48) gives $q=\frac{3}{2}$. The albedo calculated in this way is generally called the illustrative albedo and is denoted as A_i. For a body whose surface is not orthotropic, the physical meaning of A_i is as follows: it is equal to the ratio of the actual brightness of the planet for $\psi = 0$ to the brightness which would exist if the entire surface of the body were absolutely white. Sometimes the geometric albedo A_g is used; it is the ratio of the actual brightness of the planet for $\psi = 0$ to the brightness of an absolutely white plane disk of the same radius R, located perpendicular to the sun's rays in place of the given planet.

For an orthotropic surface, the different reflectivity characteristics are related as follows:

$$r=A=A_i=A_s=\frac{3}{2}A_g.$$

The albedos of the planets and of some satellites are given in Table 4 above.

§78. **The application of surface photometry to disks of planets and satellites.** The results of integrated photometry, expressed in terms of one of the albedos, give only some average characteristic of the whole hemisphere of a planet. More differentiated results can be obtained, however, using surface photometry, whose technique was discussed in Volume I, §140.

In the simplest case, only the distribution of brightness over the disk is studied, unit brightness being arbitrarily ascribed to some particular area on the disk. In this way it is possible to study, for example, the phenomenon of limb darkening at full phase or the brightness variation with distance from the terminator, or else to compare the brightnesses of various features with one another. These data are often of great significance and they enable us to draw some important conclusions on the nature of the surface.

Even more valuable data are provided by measurements which give the brightness coefficients of different features on the surface of the planet. Such measurements require comparisons between the brightnesses of features on the disk and the intensity of sunlight; these are based on certain special observations leading to a standardization with respect to sunlight.

The simplest standardization technique calls for comparing the apparent brightness B_p of a feature on the planetary disk with the brightness B_{sc} of a white screen. The screen is situated on the earth at a sufficient distance

from the telescope that it can be observed by the same means as the planetary disk; the screen is illuminated by direct solar rays only.

Let the transparency of the earth's atmosphere at the time of observation be p_p, and let the corresponding air mass be m_p. If the radius vector of the planet is Δ_p and the angle of incidence of solar rays on a feature is i, then we will have

$$B_p = \frac{rE_s p^{m_p}}{\pi \Delta_p^2} \cos i, \tag{51}$$

where r is the required brightness coefficient of the feature, and E_s is the solar illumination at a distance of 1 A.U. Likewise, for the brightness B_{sc} of the screen we can write

$$B_{sc} = \frac{r_{sc} E_s p_s^{m_s}}{\pi \Delta_E^2}, \tag{52}$$

where r_{sc} is the brightness factor of the screen, known from laboratory tests, p_s and m_s are the transparencies of the earth's atmosphere and the air mass for solar rays at the time of observation of the screen, and Δ_E is the earth's radius vector at that time. By dividing (51) by (52) and then solving for r, it is easy to show that

$$r = r_{sc} \frac{B_p}{B_{sc}} \frac{p_s^{m_s}}{p_p^{m_p}} \frac{\Delta_p^2}{\Delta_E^2} \sec i. \tag{53}$$

The dispersing screen used in this method may either reflect (in which case its brightness is observed on the illuminated side) or scatter (in which case the brightness is measured on the reverse side in terms of the transmitted luminous flux). In either case, the position of the screen must satisfy the following three conditions: 1) the screen must be accessible to photometric observations using the same instruments as are used in the photometry of the planet; 2) the incidence of the solar rays must be strictly normal or else at some accurately known angle i_{sc}; and 3) the scattered light of the sky and of terrestrial objects must not fall on the screen. In practice, it is very difficult to combine the first requirement with the third, and special standardizing adapters are required. Scattered light is permitted in some cases, but then the final result must be corrected correspondingly by measuring the scattered-light component separately. Various techniques and schemes related to such measurements have been developed and described by Sytinskaya.

Other methods have also been proposed for photometric standardization, such as comparisons of the disk brightness with that of the solar photosphere or with the brightnesses of extrafocal disks of stars. Descriptions of these techniques are given in specialized texts on photometry.

In some cases, the results of absolute photometric measurements of planetary disks are expressed in terms of the brightness factor or apparent albedo ϱ. This is defined as the ratio of the brightness B_p of a certain area of the disk to the brightness B_{on} of an absolutely white screen oriented normal to the solar rays and at the same distance from the sun. Obviously, r and ϱ are related by the expression

$$\varrho = r \cos i.$$

§79. Structures of planetary atmospheres. Let us assume that the thickness of a planetary atmosphere is small relative to the radius of the

celestial body. If this is true then, first, the curvature of the atmospheric layers will be negligible and each section of the atmosphere can be considered as a plane layer, while, second, the atmosphere will be maintained in a uniform field of gravity (that is, the acceleration of gravity g will be constant at all heights). If the mass of the gas above a unit surface area is M, then the pressure p_0 of the atmosphere at the surface will be

$$p_0 = Mg.$$

The theory of the structure of planetary atmospheres attempts to specify the height distribution of the mass M, or, equivalently, to specify the variations of the density ϱ, pressure p, and temperature T with the height h above the surface of the body. The relation between these three physical characteristics of gases is given by the classical equation of state:

$$p = \varrho \frac{R}{\mu} T, \tag{54}$$

where μ is the molecular weight, and R is the universal gas constant.

If we pass from a height h to a height $h+dh$, the pressure variation will equal the weight of the gas enclosed in the elementary layer dh; if calculations are made for unit area, then this weight will be $g\varrho\, dh$. Consequently, the main condition for atmospheric equilibrium has the form

$$dp = -g\varrho\, dh. \tag{55}$$

In order to integrate this equation, we must know the relation between ϱ and h, which in general is very complicated and is unknown except with respect to its overall form. Calculations are therefore usually made for certain simple hypothetical models of atmospheric structures, some of which will be considered below. We will consider, in particular, the case of constant μ (chemical composition the same at all heights), and p_0, ϱ_0, T_0 will denote the values of the corresponding quantities at the surface of the planet (that is, for $h = 0$).

1. Homogeneous atmosphere. This corresponds to $\varrho = \varrho_0 =$ = constant. Integration of equation (55) gives

$$p = p_0 - \varrho_0 gh. \tag{56}$$

For $h = H = p_0/\varrho_0 g$, this equation gives $p = 0$. The parameter H is called the height of the homogeneous atmosphere, and it is often used as a characteristic of the amount of gas on a given planet, since this height is proportional to the mass M:

$$H = \frac{M}{\varrho_0}.$$

The homogeneous-atmosphere model, in spite of its arbitrariness, is widely used to solve various problems in atmospheric physics, especially in atmospheric optics.

2. Isothermal atmosphere. This corresponds to $T = T_0 =$ constant. The density ρ decreases monotonically with height. Division of (55) by (54)

gives

$$\frac{dp}{p} = -\frac{\mu g}{RT}\,dh = -\beta h; \tag{57}$$

next, by integrating and substituting the boundary values p_0 for $h = 0$ and for h, we obtain

$$p = p_0 e^{-\beta h}, \tag{58}$$

and, from (54),

$$\varrho = \varrho_0 e^{-\beta h}, \tag{59}$$

where $\beta = \mu g/RT$ is a constant characteristic of the given atmosphere. Since β is proportional to g, the drops in pressure and density with height are more rapid for higher accelerations of gravity at a given body. The isothermal atmosphere has no definite boundary, since the pressure p and the density ϱ approach zero asymptotically.

3. Polytropic atmosphere. This model assumes a linear variation of the temperature T with the height h:

$$T = T_0 - \gamma h,$$

where $\gamma = dT/dh$ is the constant temperature lapse rate. Equation (57) here becomes

$$\frac{dp}{p} = -\frac{\mu g}{R(T_0 - \gamma h)}\,dh = \frac{\mu g}{\gamma R}\,\frac{d(T_0 - \gamma h)}{T_0 - \gamma h},$$

and integration gives

$$p = p_0\left(\frac{T_0 - \gamma h}{T_0}\right)^{\frac{\mu g}{R\gamma}}, \tag{60}$$

that is, a power dependence between p and h. Since $\gamma > 0$, the temperature lapse is more rapid than in an isothermal atmosphere. For $h = T_0/\gamma$, we have $p = 0$, and this gives the height of the polytropic atmosphere. Moreover, for $\gamma = \mu g/R$ we have $h = H$ and $\varrho = \varrho_0$. Therefore, the homogeneous atmosphere is a particular case of the polytropic atmosphere.

Any real atmosphere has a much more complicated structure than those indicated by these models. In particular, there is reason to assume that planetary atmospheres, like the earth's atmosphere, can be divided into the following layers which differ in their variations of T with h and in other physical phenomena as well: 1) a troposphere, characterized by convective mixing of gases and by a monotonic temperature lapse with height; the latter property makes it structurally close to the polytropic model; 2) a stratosphere, in which there is neither convection nor a monotonic variation of temperature with height, but where strong horizontal currents are observed; as a first approximation, the relations obtained for the isothermal model are sometimes applied to this layer; 3) an ionosphere, consisting of rarefied gas ionized by ultraviolet solar radiation; 4) an exosphere, the very rarefied outermost layer, where individual molecules can acquire velocities greater than the escape value and can leave the gas envelope of the body. This phenomenon leads to a gradual reduction of the atmosphere, which of course takes place more rapidly if the escape velocity v_e of the celestial body is lower. This dissipation of the

planetary atmosphere constitutes a very important cosmogonic process, and it explains why the smaller bodies of the solar system, the asteroids and almost all the satellites, have been stripped of their gaseous envelopes.

In addition to their component gases, planetary atmospheres also contain suspended particles of liquid or solid matter. A medium with such a structure is called an aerosol, or an aerodisperse system; its gaseous base constitutes the disperse medium, while the suspended particles are called the disperse phase of the aerosol. The latter is sometimes also called aerial plankton.

Three main ways can be indicated for the formation of natural aerosols. First, the plankton elements may form in the atmosphere itself due to the condensation of some of its gas. Aerosols which originate in this way may decay in two ways, either by sublimation (the vaporization of particles) or by precipitation onto the body surface. The formation and decay of aerosols in the atmosphere itself have a decisive influence on the atmospheric thermal conditions, since these processes are accompanied by release or absorption of the latent heat of vaporization. Second, the solid surface of a celestial body can be a source of aerial-plankton particles; for example, mineral dust particles may be lifted by the wind or ejected during volcanic eruptions. The third source of suspended particles is the interplanetary medium, mainly meteoric bodies. Apparently, the number of meteoric bodies arriving per unit atmospheric surface area is approximately the same for all the planets. The meteor-decay products constitute a finely dispersed solid material, which originates in the uppermost layers of the atmosphere and settles down very slowly toward the lower layers. The invariable presence of meteoric dust in planetary atmospheres is responsible for the fact that pure gaseous media apparently never exist in nature. Furthermore, calculations carried out by the Czech scientist Link and by others show that the amount of meteoric matter is high enough to have a noticeable influence on the optical properties of the atmosphere.

§80. Optical phenomena in planetary atmospheres. The very existence of a gaseous envelope about a celestial body, as well as the properties of such an envelope, are established by observing various optical phenomena originating in the gaseous medium or in the aerosol. The principal optical phenomena are the following.

1. Refraction. The refraction of light rays in a planetary atmosphere displaces somewhat the apparent positions of features on the disk, but this displacement is too slight to be detected during studies of the trajectories described by the projections of the features in the image plane due to rotation. The various refraction phenomena are therefore observed mainly near the limb, when the disk occults a star or some other celestial body. Then, the stellar image is displaced due to refraction of its rays, and there is a gradual variation of its apparent brightness before disappearance and after emergence due to the refractive attenuation connected with the transformation of a parallel beam into a convergent one.

2. Scattering of light. The application of scattering data to the earth's atmosphere has been discussed in Volume I, Chapter XXIX. When a planet is observed from outside, the scattered light shows up as an atmospheric haze, whose brightness adds to that created by the light reflected from the solid body surface; this haze weakens the brightness contrasts between various surface features. When the optical thickness of the atmosphere is considerable, the surface becomes entirely invisible. For

comparatively transparent atmospheres, the scattering effect is enhanced in those regions where the gas turbidity is high due to large admixtures of suspended particles; these regions therefore appear to be comparatively light and opaque formations which form clouds and mists. The propagation of scattered light beyond the terminator is responsible for twilight phenomena in the night hemisphere of a planet.

3. Absorption of radiant energy. For the gaseous component, the absorption is quite selective, producing the discrete absorption bands of the atmospheric gases in the spectrum of the radiant flux reflected by the planet. These bands make possible spectroscopic studies of the chemical composition of planetary atmospheres. Such studies are easy only when the absorption bands are sufficiently strong; otherwise, the detection and identification of weak planetary bands among the lines of the solar spectrum is a difficult task. The identification of gases which are also contained in the earth's atmosphere is further complicated by the fact that the absorption bands of the planetary atmosphere are superimposed onto telluric lines, due to absorption in the earth's atmosphere. In this case, observations should be made at times when the line-of-sight velocity of the planet relative to the earth is sufficiently high that lines of planetary and telluric origin can be separated as a result of the Doppler effect.

Among other optical phenomena, we should mention the influence which light scattered in the atmosphere has on the polarization curve and also gas-luminescence phenomena analogous to polar auroras on the earth, the existence of which on other planets may be assumed.

Optically, planetary atmospheres are characterized by various parameters, the principal ones being: 1) the angle of horizontal refraction ω; 2) the atmospheric transparency p, defined as the fraction of the radiant flux transmitted by the atmospheric layer for vertical incidence of the light, or else the optical thickness τ, which is related to p by the expression $\tau=-\ln p$; 3) the scattering indicatrix, which is a vector diagram representing the angular distribution of the scattered radiant flux (see Volume I, §184); and 4) the lightness factor η, equal to the ratio $\tau_s/(\tau_s+\tau_a)$; here τ_s is the optical-thickness component due to scattering, $\tau=\tau_s+\tau_a$ is the total optical thickness, and τ_a is the component due to absorption. All these parameters vary in general with the wavelength λ. Some of them, such as the scattering indicatrix and the factor η, may also vary with the height above the surface.

The photometric theory of a planetary atmosphere seeks to establish equations expressing the distribution of apparent brightness over the disk for for any phase, on the basis of known values of the above parameters and using data on the reflection of light by the surface. This theory must also solve the reverse problem, namely, given the brightness distribution over the disk, to find the optical characteristics of the atmosphere and the surface of the planet. The solution of this problem in any kind of general form is one of the very complicated tasks of theoretical physics; it has been attempted by the German scientists Lommel, von Seeliger, and Schönberg, by King in England, and by Khvol'son and Fesenkov in Russia. Considerable progress was made recently due to the outstanding work of Ambartsumyan in the USSR and Chandrasekhar in the USA. However, at present the brightness can be calculated in practice only for certain simplified models, for instance, for models which neglect the refraction and curvature of atmospheric layers, with fairly simple indicatrices.

The problem is rendered more difficult by the need to take into consideration both the effect of self-illumination (that is, of higher-order scattering in the atmosphere) and also radiant exchange between the atmosphere and the planetary surface, it being necessary to consider polarization effects in all cases. To simplify the actual situation, we may say that the brightness at any point of a disk enveloped in an atmosphere is made up of the following components: 1) the brightness of the solid surface, produced by reflection of the directional beam of solar rays (§76), the luminous flux being attenuated by the atmosphere twice (on its way from the sun to the surface and then from the surface to the observer); 2) the surface brightness caused by illumination from scattered atmospheric light (that is, from sky light); 3) the brightness of the haze produced by primary scattering of sunlight in the atmosphere; 4) the brightness of the haze produced by higher-order scattering (a ray reflected by an element of the medium encounters another element and is again reflected, etc.); 5) the brightness of the haze produced by the scattering in the atmosphere of the luminous flux reflected from the solid surface. The relative significance of the various components will differ in different cases; here, it will be useful to consider the following three cases.

1. The optical thickness of the atmosphere is virtually infinite, so that the light reflected by the body surface has no noticeable influence on the disk brightness. In this case, the brightness results only from the light scattered in the atmosphere.

2. The optical thickness of the atmosphere, though high, is still finite. In this case, direct rays reflected by the body surface are not transmitted through the atmosphere, and the surface features remain invisible. However, the illumination of the atmosphere by light reflected from the surface and then repeatedly scattered above it influences the luminous regime of the upper, as well as the lower, atmospheric layers. Consequently, the brightness distribution over the disk depends on the surface albedo.

3. The optical thickness of the atmosphere is low. In this case the surface is visible through the atmosphere. The brightness at each point of the disk is the sum of the surface brightness, attenuated by extinction in the overlying atmospheric layer, plus the brightness of the atmospherically scattered light.

Planets with cloudy atmospheres, such as Venus, Jupiter, Saturn, Uranus, and Neptune, must be classified either in the first or second group; the only example of the third type is Mars.

§81. Temperature conditions. The temperatures of the body surface and the gaseous envelope are primarily determined by the distance from the sun, since the intensity of solar irradiation varies in inverse proportion to the square of this distance. The temperature distribution over a celestial body can be studied either theoretically or by observation.

The simplest example of a theoretical analysis of planetary temperature conditions is the calculation of the equilibrium temperature. This is the temperature attained by a surface which is under constant irradiation and loses heat only by emission into the surrounding space. At this temperature, the influx of radiant energy (from sunlight) will be equal to the radiation losses, that is, the incoming and outgoing energy fluxes will be in equilibrium.

Let us consider a unit surface area located in a vacuum; the surface lies perpendicular to the solar rays at a distance Δ from the sun. In unit time

it receives an amount of radiant energy equal to E_S/Δ^2, where E_S is the solar constant and Δ is expressed in astronomical units. It is assumed that the irradiated surface is a perfect blackbody, so that all the incident energy is absorbed, and that the absorbed energy is lost only by radiation into the surrounding space. In this case, the radiation losses per unit surface area per unit time are given by Stefan's law; they are equal to σT_1^4, where T_1 is the absolute equilibrium temperature. The condition of equilibrium gives

$$\sigma T_1^4 = \frac{E_S}{\Delta^2},$$

from which we can find the equilibrium temperature T_1:

$$T_1 = \sqrt[4]{\frac{E_S}{\sigma\Delta^2}} = \frac{392^\circ}{\sqrt{\Delta}}. \qquad (61)$$

In practice, conditions not widely divergent from those assumed here as a basis for calculation exist at the subsolar point of a low-albedo celestial body without an atmosphere, for instance, at the center of the illuminated hemisphere of the moon or Mercury.

Next let us consider an alternative model. We consider a perfectly black spherical body of radius R located a distance Δ from the sun. Per unit time its illuminated hemisphere receives on the whole an amount of radiant energy equal to $\pi R^2 E_S/\Delta^2$. It is assumed that the temperature at all points of the body surface is constant and equal to T_2, that is, the thermal energy is distributed uniformly over the entire body surface, from which it is radiated. The equilibrium condition is now

$$4\pi R^2 \sigma T_2^4 = \pi R^2 \frac{E_S}{\Delta^2},$$

from which

$$T_2 = \sqrt[4]{\frac{E_S}{4\sigma\Delta^2}} = \frac{T_1}{\sqrt{2}} = \frac{277^\circ}{\sqrt{\Delta}}. \qquad (62)$$

In most cases, the actual surface temperature of the illuminated hemisphere of a planet will lie between T_1 and T_2; the corresponding values are given in Table 5.

TABLE 5

Planet	Solar irradiation at the mean distance			Ratio of irradiations at perihelion and aphelion	Surface temperature, °C		
	relative (earth = 1)	absolute			equilibrium		measured radiometrically
		thermal, $\frac{g\cdot cal}{cm^2\cdot min}$	luminous, 10^3 lux		T_1	T_2	
Mercury	6.68	12.7	901	2.31	+358	+172	+400
Venus	1.91	3 6	258	1.03	+191	+54	−40
Earth	1.00	1.9	135	1.07	+119	+4	—
Mars	0.431	0.81	58	1.45	+43	−51	+20 −60
Jupiter	0.0370	0.070	5,0	1.21	−100	−151	−140
Saturn	0.0110	0.021	1,48	1.25	−145	−183	−150
Uranus	0.0027	0.0052	0,366	1.21	−184	−210	−200
Neptune	0.0011	0.0021	0,149	1.04	−201	−222	−200
Pluto	0.0006	0.0012	0,087	2.77	−211	−229	—

A further development of the theory will take into consideration, on one hand, the angle of incidence of solar rays at the surface, and, on the other hand, the divergence of the actual surface properties from those of a perfect blackbody. It should be noted that most of the solar energy is concentrated in the visible and near-infrared spectral regions, for which the albedo gives the amount of absorbed energy. However, the radiated energy is concentrated in the far infrared, and the corresponding albedo specifies the emissivity of the actual surfaces. The theory behind such calculations is the subject of mathematical climatology. The more complex models considered in this field take into consideration the influence of the atmosphere on the temperature regime of the planet. The gaseous envelope, on one hand, absorbs part of the solar energy, and a certain fraction of this heat is expended in such processes as the evaporation of water, the formation or decay of aerosols, the setting up of atmospheric currents, etc.; on the other hand, heat transfer by atmospheric winds promotes a uniform distribution of the heat and creates conditions approximating those assumed in the T_2-model. The atmosphere may also raise the surface temperature because of the so-called greenhouse effect. The latter arises due to the fact that the absorption spectra of some gases (for example, of CO_2 and water vapor) consist of bands located predominantly in the far infrared. Consequently, the solar flux, whose wavelengths, as we have indicated, lie mainly in the visible spectrum, is only slightly absorbed in the atmosphere, while the long-wave flux emitted by the surface experiences strong absorption, also leading to a certain accumulation of heat.

Some very important results in the field of mathematical climatology, relating both to the earth and to other planets, were published by the Yugoslav scientist Milanković. Planetary temperatures are experimentally determined by measuring with a thermocouple the energy received from a certain surface area (see Volume I, §165). The experimental temperatures are also listed in Table 5.

§82. Planetary radio astronomy. Radio-astronomical observations of planets and satellites can be carried out in two ways: 1) by measuring the planetary emission at the corresponding frequencies, and 2) by radio probing, that is, by sending a beam of radio waves from the earth and observing the echo (the signal reflected from the surface of the body).

Planetary radio emission was first discovered for Jupiter (the first indications were obtained as early as 1950), and then for the moon and some other planets. The wavelength range used in these observations lies between 8 mm and 75 cm. In many cases the emission is very weak, so that the measurements require large radio telescopes and highly accurate measurement techniques. In the USSR, observations of this kind have been developed extensively at the Main Astronomical Observatory (Pulkovo) of the USSR Academy of Sciences (Khaikin and Kaidanovskii), at the Lebedev Institute of Physics (Moscow) of the Academy of Sciences, at the Crimean branch of this institute (Kuz'min and Salomonovich), and at the Gor'kii Radiophysical Institute (Troitskii and co-workers).

Planetary radio waves can be either thermal or nonthermal in origin. Thermal radio emission does not differ in principle from the radiation observed by optical means, and its intensity is expressed by ordinary emission laws. In particular, the spectral distribution of blackbody radiation is expressed by Planck's formula, and for the large wavelengths λ

encountered in radio astronomy this reduces to the Rayleigh-Jeans equation

$$S_\lambda = \frac{2kT\omega}{\lambda^2}, \tag{63}$$

where S_λ is the flux density per unit area and unit frequency band, ω is the solid angle subtended by the disk of the body at the observation point, T is the absolute temperature, and k is a constant. Radio-astronomical observations give the values of S_λ, and then the temperature T is calculated from (63). The temperature obtained in this way is generally called the radio-brightness temperature of the body; it approaches the actual temperature to the same extent that the properties of the body surface approach those of a blackbody. Thermal emission has so far been observed for Mercury, Venus, the moon, Jupiter, and Saturn, and the values found for T are close to those obtained from radiometric observations in the infrared region (8 to 13 μ).

When observations of this kind are interpreted, it must be remembered that the surface material of a planet is partially transparent at radio frequencies, so that the radiation detected originates not only at the outermost surface but also at lower-lying strata. The radiometric values of T therefore represent average temperatures for a layer whose thickness increases with the wavelength λ. Consequently, a comparison of measurements for various λ makes it possible to obtain the temperature gradient for the external ground layer; so far, this has been done only for the moon.

Nonthermal radiation typically has a frequency distribution which does not correspond to formula (63). Such radiation has been discovered for Jupiter, and, apparently, for Venus. Its origin is related to processes occurring in the ionosphere of the planet or in the surrounding radiation belts.

Radar methods (radio probing) were first applied to the moon in 1946. Then, in 1961, in the USSR, the USA, and England, radar soundings were made of Venus. The latter were not only important for planetary physics but also made it possible to determine with higher accuracy the astronomical unit of length. Previous determinations of this quantity, based on measurements of the solar parallax, had relative errors of the order of 10^{-3} of the measured value, while the new radar data have errors of the order of 10^{-5}. In 1962 Mercury and Mars were also radio probed, and in 1963 Jupiter was probed (Kotel'nikov and co-workers).

The further application of radar methods is limited by the weakness of the reflected signal, whose intensity is inversely proportional to the fourth power of the distance from the object. However, by sending signals of considerably higher intensity and by improving the sensitivity of the receiving equipment, further progress in this field will be possible.

The signal reflected from a planet gives certain data on the nature of the celestial body. For example, by comparing the power of the transmitted and received signals, we can find the radio-frequency albedo of the planet. Since fog and clouds are transparent to radio waves, this value will represent the surface albedo of planets whose solid surfaces are surrounded by a cloudy atmosphere and thus are inaccessible to optical measurements. If different surface regions of a planet have different albedos, the rotation of the planet will show up as a periodic variation of the mean albedo, and this will give valuable information on the nature of the planet.

There is also another method for studying planetary rotation by means of radar. Let us assume that the transmitted signal is monochromatic, that is, it corresponds to a very narrow spectral band of width $\Delta\lambda$. When the planet rotates, the Doppler effect arising at the limb changes the wavelength, and the reflected signal covers a wider band in the spectrum; the rotation period can then be determined from the difference in band widths (see §72).

If signals are transmitted in the form of very short pulses, then the reflected signals will be stretched out in time somewhat. This takes place because different zones of the planet surface are located at different distances from the earth (the marginal zones of the disk are farther away by a distance equal to the planetary radius), so that the transit time for the signals will be different. By analyzing this effect, we can determine the size and shape of the solid body of the planet, and we can also draw certain conclusions on the nature of its surface (whether it is mat, pitted, or specular).

Radar techniques can also be applied to solve other problems in the physics of planets, for instance, to investigate planetary ionospheres. *

* [See reference list following Chapter XII.]

V. V. SHARONOV

N 67-15496

Chapter XII

A DESCRIPTION OF INDIVIDUAL PLANETS

§83. Mercury. The conditions for observing this planet are unfavorable because of its proximity to the sun and the smallness of its angular dimensions. Visual observations have detected blurred spots on the disk, located at about the same positions relative to the terminator for equal elongations, but with positions that vary with elongation. This indicates that the period of axial rotation for Mercury is equal to its period of revolution about the sun, so that Mercury always has the same side facing the sun. The considerable orbital eccentricity produces libration, resulting in an alternation of night and day with a period of 88 days in a certain zone near the terminator.

Many observers have charted the surface of Mercury from the results of visual observations; in the years 1942—1944 numerous photographs were taken at the Pic du Midi Observatory, and these served as a basis for the first photographic map. In such maps it is assumed that the equator of Mercury is coplanar with its orbit, and that the initial meridian passes through the subsolar point at the time of perihelion of the planet. A comparison of charts drawn up by different authors reveals only a remote likeness as far as the major spots are concerned and shows wide discrepancies for the fine details.

A comparison of recent charts with those compiled in the 19th century led Dollfus to conclude that the periods of rotation and revolution coincide to at least within 1/10,000, but that the agreement with observations is improved if the equator is assumed to make an angle of 7° with the orbit.

Repeated careful observations have been made, attempting to detect traces of an atmosphere during the transits of Mercury over the solar disk (Lomonosov effect) or during lower conjunctions with the sun (lengthening of the cusps); however, these gave negative results. It has thus been established that the gaseous envelope of the planet, if it exists at all, has a density not exceeding 1/1000 of the density of the earth's atmosphere.

Surface photometry has not yet been applied to the disk of Mercury. The integrated-light photometry carried out by Müller in the 19th century gave two expressions for the variation of the magnitude m with phase, and both of these represent almost equally well the observation results (L = 1 A.U., Δ equals the mean distance):

$$\text{I. } m = -1.04 + 0.0368\,(\psi - 50^\circ).$$
$$\text{II. } m = -0.90 + 0.0284\,(\psi - 50^\circ) + 0.000102\,(\psi - 50^\circ)^2.$$

According to Russell's calculations, these formulas correspond to the following values of the various photometric elements:

Formula	g	A_g	q	A_s
I	−0.88	0.164	0.42	0.069
II	−0.06	0.077	0.74	0.055

Here g is the absolute magnitude, A_g is the geometric albedo (for an angular disk semidiameter at unit distance of 3".45), q is the phase integral, and A_s is the visual spherical albedo. Later, Danjon used a new series of observations to obtain

$$m = -0.21 + 3.80\,\frac{\psi^\circ}{100} - 2.73\left(\frac{\psi^\circ}{100}\right)^2 + 2.00\left(\frac{\psi^\circ}{100}\right)^3,$$

which also corresponds to a spherical albedo of 0.055.

Photometrically, Mercury has a certain similarity to the moon, and this may indicate that the two bodies have similar surface structures. This assumption is also supported by Lyot's polarimetric observations, according to which the curve of phase vs. polarization for Mercury is close to that of the moon (see Figure 105). We thus conclude that the surface of Mercury is a mountainous, very furrowed, stony desert. However, all attempts to detect some traces of mountains during careful visual observations of the terminator of Mercury have failed.

The equilibrium temperature T_1 for Mercury is +358°C, while radiometric measurements give a value close to +400°C; these figures provide an indirect proof of the slowness of the axial rotation.

§84. Venus. We see from Tables 1 and 3 of Chapter XI that the size, density, and mass of Venus are close to those of the earth. Since Venus is an inferior planet, we can plot its phase curve almost completely and we can find all the necessary photometric parameters. The variation of the stellar magnitude m_0 of Venus, reduced to the mean distance of the planet from the sun and to a distance of 1 A.U. from the observer, can be plotted as a function of the phase angle ψ according to Müller's visual observations. Here, the empirical formula

$$m_0 = -4.71 + 0.01322\psi + 0.000000425\psi^3,$$

corresponding to a phase integral q = 1.20 and a spherical albedo of 0.59, is used. It follows from this that the surface color of Venus is very light, and more recent measurements have given the same result.

Previous observers have disagreed concerning the color of Venus. Direct comparisons with the sun made during the last decades demonstrated that the light of the planet is slightly yellower than the sun, the difference not exceeding $0^m.2$ of color index. Visual telescopic observations, even for favorable conditions (large angular size, high disk brightness), detect only a few, in many cases unreal, features; these show up as blurred spots, relatively light areas near the cusps, and indentations in the terminator. Therefore, a characteristic property of the Venusian surface is its exceptionally uniform light color and the absence of visually distinguishable features. Correspondingly, nothing definite is known concerning the rotation of Venus about its axis. Doppler-shift measurements have also failed, since lines at the limb show no appreciable displacement relative to the center of the disk. The first radar determinations of the Venusian rotation gave a long period of over 100 days, a value close to the period of revolution about the sun.

It is an accepted fact that the apparent aspect of Venus is conditioned by a continuous cloud cover hiding the solid surface of the planet. Attempts to penetrate this cloud envelope by observations in infrared light have failed, since infrared photographs of Venus show the same aspect as those in visual light. On the other hand, photographs taken with ultraviolet rays show easily distinguishable features, mainly in the form of blurred dark bands perpendicular to the line of the cusps. Kuiper has suggested that the boundaries of these bands follow the planetographic parallels, and on this basis he attempted to determine the equatorial elements of Venus. For the north-pole coordinates he found $\alpha = 53°$ and $\delta = +81°$, which gives an equatorial inclination of 32° to the plane of the orbit. Later, Richardson used more comprehensive material to obtain $\alpha = 311°$ and $\delta = +64°$, giving an inclination of 14°.

The existence of a Venusian atmosphere was first established precisely by the famous observations of Lomonosov, made in St. Petersburg during the transit of Venus across the sun on 26 May (6 June according to the Gregorian calendar) 1761. At the initial phase of the transit, Lomonosov noticed "a glow, as fine as a hair" encircling the limb of Venus prior to its superposition onto the solar disk; an analogous phenomenon was observed as Venus left the solar disk. Lomonosov supplied the correct explanation of this phenomenon. He wrote: "This shows that the solar rays are refracted in Venus's atmosphere." He concluded that "the planet Venus is surrounded by a notable air atmosphere, like (if not larger than) that surrounding the surface of our own globe."

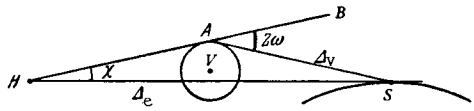


FIGURE 121

The "Lomonosov effect," namely the bright fringe encircling the limb of Venus outside of the solar disk, was observed during three subsequent transits, in 1769, 1874, and 1882. The fringe appears in the form of whiskers, sprouting from the solar limb and merging at a certain time into a continuous arc; this fringe is formed by the image of the solar limb refracted in the Venusian atmosphere. The corresponding position V of Venus is shown in Figure 121, where H is the point of observation, and S is the point on the solar limb where the ray SA originates; this ray, after being refracted by an angle $BAS = 2\omega$ in the atmosphere of Venus, passes through the point H. Let us now denote as $\chi = \angle BHS$ the angular distance of the outermost point on the Venusian limb from the edges of the solar disk at the time when the whiskers merge. Then, since the sides HS and AS of the triangle HAS can with sufficient accuracy be taken equal to the radius vectors Δ_e and Δ_v of the earth and Venus, we obtain

$$\sin 2\omega = \frac{\Delta_e}{\Delta_v} \sin \chi$$

or, the angles ω and χ being small,

$$\omega'' = \frac{\Delta_e}{2\Delta_v} \chi''. \qquad (1)$$

We now can easily find the horizontal refraction ω for the transparent layer of the Venusian atmosphere. For instance, the application of formula (1) to the results of the two transits in the 19th century gave a value of ω close to 20". The more rigorous theory of the Czech astronomer Link gives a higher value of ω for the same data; his value is close to 40", and occasionally as high as 1'.7, whereas the horizontal refraction in the earth's atmosphere is 35'.

The occultations of stars by Venus are of considerable interest with respect to atmospheric investigations of this planet. In particular, the numerous observations of the occultation of Regulus made in various countries on 7 July 1959 have been analyzed by Martynov in the USSR, and also by de Vaucouleurs and Menzel. The following results were obtained: the atmospheric layer in which the refractive attenuation of light makes the star vanish lies at a height of about 70 km above the apparent cloud surface. The pressure in this layer is about 2.6 dyne/cm^2, and the total mass of the overlying gas is $3 \cdot 10^{-3}$ g/cm^2.

Another phenomenon closely associated with the atmosphere of Venus is the so-called "lengthening of the cusps," discovered at the end of the 18th century by Schröter and W. Herschel. During the three phases when Venus has the form of a fairly thin crescent, the light arc along the limb covers more than 180°. The angle of lengthening U is the angle between a disk radius drawn through the cusp of the crescent and the perpendicular to the intensity equator.

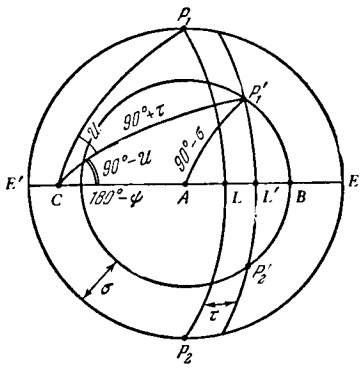


FIGURE 122

The lengthening of the cusps is caused by the following fact: due to twilight phenomena, the apparent terminator of Venus is displaced toward the dark hemisphere by an amount equal to the twilight arc σ. In order to derive a relation between U and σ, let us consider Figure 122. In this figure, P_1EP_2E is the great circle corresponding to the position of the terminator in the absence of an atmosphere, for an observer facing the night hemisphere of Venus, A is the point diametrically opposite to the pole of illumination (for which the sun is at the nadir), C is the center of the disk, P_1LP_2 is the orthographic limb, EAE' is the intensity equator, $\smile CA = 180° - \psi$, where ψ is the phase angle, and the circle $P_1'BP_2'$ is the apparent terminator displaced toward the dark hemisphere by an arc σ of a great circle. Let us assume that the apparent limb is displaced by refraction through an arc τ toward the invisible hemisphere. The apparent cusps are then the points P_1' and P_2'; the dark arc of the limb $P_1'L'P_2'$ enclosed between these apparent cusps is then less than 180°, and the lengthening of one of the cusps is represented in the drawing by the spherical angle $P_1'CP_1$.

In the spherical triangle $P_1'AC$ we have: $\smile CA = \varphi = 180° - \psi$, $CP_1' = 90° + \tau$, and $AP_1' = 90° - \sigma$. For this triangle we may write

$$\sin\sigma = \cos\psi \sin\tau + \sin\psi \cos\tau \sin U. \qquad (2)$$

Some authors, neglecting twilight phenomena, have attributed the displacement of the limb and terminator exclusively to atmospheric refraction. In this case, we must set $\sigma = \tau = \omega$, so that formula (2) gives

$$\operatorname{tg} \omega = \frac{\sin \psi \sin U}{1 - \cos \psi}.$$

Calculations using this formula give very high values of ω (2 or 3°), and they have led to the suggestion that the atmosphere of Venus is exceptionally dense. However, Russell, and subsequently other authors as well, have shown that such high refraction cannot in general be observed, since it implies an atmospheric layer of such a high optical thickness that, due to molecular scattering, the atmosphere will be entirely opaque, and the refracted beam will not reach the observer. Consequently, many authors now neglect the influence of refraction (this can be partly justified by the data for the Lomonosov effect), and they attribute the lengthening of the cusps to twilight phenomena alone. By setting $\tau = 0$, we find from (2) that

$$\sin \sigma = \sin \psi \sin U$$

or, by taking $\varphi = 180° - \psi$ and seeing that the angles are small, we obtain

$$\sigma' = \varphi' \sin U. \qquad (3)$$

Measurements of U made against the background of the daylight sky give $\sigma = 3°$, and measurements at night give values of σ up tc 7°. This indicates a considerable development of twilight phenomena on Venus.

But how is it possible to reconcile the low refraction indicated by observations of the Lomonosov effect with these well-developed twilight phenomena ? With respect to this, Sharonov advanced the hypothesis that the cloud cover constituting the apparent surface of Venus, being situated at a high altitude in the atmosphere, is thin and semitransparent. The refraction observed during the transits of Venus originates in the transparent gaseous layer underlying the cloud cover, while the twilight phenomena are caused by scattering in the transparent gaseous layer above the cloud cover.

An analysis of the phase curve of Venus was made by Sobolev. This investigator used a simplified theory of higher-order scattering, together with the surface-photometry data of Barabashov and his co-workers, and also the polarimetric observations of Venus and of terrestrial cloud forms made by the French astronomer Lyot. This analysis indicated that the scattering layer of Venus must be composed of comparatively coarse elements, not unlike the water droplets constituting terrestrial clouds. This problem, however, is a difficult one and, in particular, it is impossible to establish whether the scattering elements are liquid spherical drops, snowlike crystals, or irregular, solid, dustlike particles. Consequently, the nature of the atmospheric aerosol on Venus remains a mystery.

Spectroscopic investigations of the Venusian atmosphere failed to give any definite results for a long time, since no planetary absorption bands could be detected. It was only in 1932 that Adams and Dunham (Mount Wilson Observatory) found absorption bands in the near infrared of the spectrum of Venus, and these were subsequently identified with carbon dioxide. So far, up to 15 bands have been identified. In order to estimate the CO_2

content in the planetary atmosphere, careful comparisons were made between the intensities of all these absorption bands and CO_2 spectra obtained under laboratory conditions. The results, however, are somewhat contradictory, and they give CO_2 contents from 100 to 1000 m of a standard atmosphere (pressure 760 mm, temperature 0°C). The reason for these discrepancies is that the line intensity for a given line-of-sight amount of gas depends on both the pressure and the temperature, neither of which is known for the atmosphere of Venus. Moreover, observations of laboratory lines and observations of lines in a planetary atmosphere (illuminated by sunlight and characterized by substantial scattering) correspond to entirely different conditions.

Recently, several reports have been published on the discovery of weak bands of other gases in the spectrum of Venus. For example, Prokof'ev studied the telluric bands of oxygen in spectrograms of Venus taken at the Crimean Astrophysical Observatory using high-dispersion equipment. He discovered a certain asymmetry in the profiles of the spectral lines, which could be attributed to the superposition of weak oxygen lines of the planetary atmosphere shifted by the Doppler effect. After processing the spectroscopic observations made by Ross and Moore from a high-altitude balloon, Strong concluded that the 1.13-μ water-vapor band is enhanced somewhat in the spectrum of Venus; this can be ascribed to the presence of 19μ of precipitated (liquid) water in the relevant atmospheric layers. Dollfus came to an analogous conclusion after his comparison (winter 1963) of the spectra of Venus and the moon under high-mountain conditions; analyzing the 1.4-μ absorption band, he found a content of 10^{-2} g/cm^2 of H_2O vapor in the planetary atmosphere. All these results, however, require further verification.

In 1953, Kozyrev (Crimean Astrophysical Observatory) reported that he had observed signs of an emission spectrum on the dark side of Venus, the bands of which could be partly identified as molecular nitrogen. However, analogous observations carried out at other observatories have failed to confirm this result, and Kozyrev himself subsequently could not detect this atmospheric emission again either. At present, the existence in the Venusian atmosphere of such emission detectable by an earthbound observer still remains undecided.

The equilibrium temperatures of a perfectly black surface for the mean distance of Venus from the sun are $T_1 = +191$ °C and $T_2 = +54$ °C, but a surface in the form of a cloud layer may in reality have an entirely different temperature. Previous works (1924—1926) on radiometric determinations of this temperature, which gave −23 to 0°C for the dark hemisphere and +25 to 50° (and even up to +60°C) for the illuminated hemisphere, are now regarded as erroneous even by the authors themselves. According to recent data, the temperature of the apparent surface of Venus is close to −40°C for both night and day, since hardly any diurnal temperature variation could be observed. Sinton, in 1953, obtained −39°C for the day hemisphere, and Pettit and Nicholson in 1955 published a new analysis of their 1923—1928 measurements, giving −37°C for the night side and −42°C for the day side; since the sign of the difference is obviously opposite to that expected, this difference should be regarded as unreal.

Another means of measuring the atmospheric temperature of Venus is an analysis of the fine rotational structure of the CO_2 absorption bands. This

gives values of the order of −10°C. Some very important data on the temperature regime of Venus have been provided by radio-astronomical observations. Measurements of the planetary heat on 8-mm waves give radio-brightness temperatures up to +80°C; measurements at various frequencies in the range from 3 to 21 cm give identical, but unexpectedly high, temperatures, ranging from 300 to 350°C.

The considerable scatter of the temperature values obtained is quite natural, since different methods of measurement penetrate to different depths in the planetary atmosphere; thus, the temperatures correspond to different atmospheric levels.

Many hypotheses have been advanced concerning the nature of Venus, the basic assumptions being adapted to the accumulation of new facts. The once popular hypothesis of a moist tropical climate, analogous to that prevailing on the earth during the Carboniferous period, and also the later suggestion that Venus is entirely covered with water, separating the solid surface from the atmosphere, are difficult to reconcile with the negligible water-vapor content in the atmosphere, according to spectroscopic data. At present (summer 1963), the following three models of the planetary atmosphere are subjects of lively discussion.

1. The greenhouse model. According to this model, the temperature of +320°C indicated by radio-astronomical observations in the 3-cm to 21-cm wavelength range represents the true temperature of the surface and the adjoining atmospheric layer. Such a high temperature is sustained by the greenhouse effect resulting from the exceedingly thick atmosphere. Since carbon-dioxide gas cannot supply the required effect, we must assume the content of water vapor in the atmospheric layer to be 10 g/cm^3.

Sagan's calculations give the following results: the atmospheric pressure at the surface is 2 to 5 atm higher; the atmospheric layer up to a height of 36 km corresponds to the adiabatic model with a temperature gradient of 10° per km. At the boundary of this layer the pressure is close to 0.1 atm, and the temperature corresponds to the value of −40°C obtained from radiometric observations. Under these conditions, the water vapor becomes saturated, and a continuous cloud cover is formed at this altitude, which is similar to terrestrial cirrus clouds composed of ice crystals. The atmospheric layer between 36 and 105 km is transparent and isothermal, with a temperature of −40°C throughout. The altitude of 105 km corresponds to the layer in which stars disappear during occultation; the pressure at this altitude is $2 \cdot 10^{-6}$ atm. It is assumed that approximately at this level another thin haze layer exists, which is transparent to visible light but which absorbs ultraviolet rays. The structure of this layer shows up in the form of dark formations on ultraviolet photographs. Possibly, the constituent particles are polymers of carbon suboxide $(C_3O_2)_n$.

2. The eolospheric model. This model was proposed by Öpik. Here it is assumed that an adiabatic layer 15 km thick adjoining the surface constitutes the eolosphere, namely a zone with a high dust content and characterized by exceptionally strong atmospheric currents. The latter are related to a convective transfer of heat from the illuminated hemisphere to the night side of the planet, which always turns the same side to the sun. The high friction generated by these currents keeps the temperature above 300°C. The surface pressure is close to 4 atm. The pressure at the boundary of the eolospheric layer is 0.6 atm, and the temperature there is

−10°C (the value obtained from the rotational structure of the CO_2 bands). A thick cloud layer consisting of mineral dust particles exists at this height. At an altitude of 36 km there is a second haze layer; the pressure there is 0.08 atm and the temperature is −40°C. This model gives a constant temperature for all phases, but this is true only for the radiometric temperatures and is inconsistent with the observations at radio frequencies.

3. The ionospheric model. This model is based on the assumption that the radio waves emitted by Venus in the 3 to 21 cm range are not thermal: they are emitted not by the planetary surface but by the ionosphere, due to electron transitions. However, to explain the observed radio intensity, a very thick ionosphere must be assumed; in addition, the ionospheric electron concentration must have the improbably high value of about 10^9 electrons per cm^3, which in turn requires the introduction of several additional hypotheses. Moreover, such an ionosphere would completely absorb the radio waves used in radar techniques, and no reflected signal would be observed; it is difficult to reconcile this with the successful probing of Venus by means of radar. Finally, according to this model the atmospheric pressure at the planetary surface is a mere 0.3 to 1 atm, and the temperature is close to +80°C, the value obtained on 8-mm waves (for which the ionosphere is transparent). At an altitude of 15 km there is a cloud layer of unknown nature which is opaque in the visible spectrum.

Combined hypotheses have also been advanced, which assume that the radio emission of Venus slowly varying with the phase is thermal, while the rapid irregular fluctuations noted by some observers are due to ionospheric processes. Large-scale radio-astronomical observations of Venus are rapidly filling in the blanks in our knowledge of this planet; even more promising are investigation methods which will utilize astronautics. The models and schemes claiming to describe the physical nature of Venus will therefore be revised continually as new data are amassed.

§85. The earth. Geometrical, mechanical, and physical studies of the earth are the subjects of special sciences, such as geodesy, geophysics, geography, and geology. Therefore, here only a few problems pertaining to the nature of our planet will be discussed.

The brightness of the earth can be found by photometric measurements of earthlight on the moon. The brightness of limb points lying on the intensity equator are measured to give the brightness ratio for continents illuminated by the reflected light of the earth and the direct light of the sun. Then, given the phase curve of lunar-limb brightness, we can find the ratio of the earthlight to the brightness of the continents at full moon, which directly establishes the ratio of the earth's brightness at a particular phase to the brightness of the sun. We can thus plot a large portion of the earth's phase curve. Finally, by extrapolating the curve to 0° phase angle, we can find the brightness of the earth at full phase, and then by applying the formulas of §77 we can calculate the various albedos.

Considerable difficulty is encountered here due to the overlapping (onto the dark part of the disk) of the light of the crescent, which is scattered in the earth's atmosphere and in the optics of the instruments. The most reliable results are those obtained by the French astronomer Danjon, using a special visual photometer of his own construction. His brightness curve

is described by the following data:

ψ	0°	30°	60°	90°	120°	150°
m	−4.57	−4.09	−3.53	−2.79	−1.82	−0.51

The values of m have been reduced to distances of 1 A.U. from the sun and from the observer. The color of the earthlight is bluish, as was established by Tikhov back in 1914. According to Markov the earth's color is in the spectral class FO, while Orlova has given a color index of $-0^m.27$ for the earth.

The earth's disk as viewed from space should first of all be a variable combination of white spots, corresponding to cloudy regions. The earth's surface will show between the clouds, but the visibility of surface features will be lowered considerably by atmospheric air glow, particularly near the edges of the disk.

These conclusions based on the results of earthbound observations and theoretical arguments were fully corroborated by the observations of the first astronauts, Yurii Gagarin and German Titov, from space. In particular, the astronauts took excellent color photographs of the earth; these show clearly both the white clouded areas and the blue interstices between them, where the color is determined by Rayleigh scattering in the earth's atmosphere (this mechanism also gives the above-mentioned blue tint to the integrated light of the earth).

A lunar eclipse is another astronomical phenomenon involving the earth's atmosphere. Kepler observed that the reddish illumination of the lunar disk in the earth's umbra is due to refraction; for rays tangent to the earth's surface the refraction is twice the horizontal refraction: $2\omega = 70'$. These refracted rays form a cone with its vertex lying at one third of the distance to the moon. The entire umbral cone is therefore illuminated by the refracted rays, and the selective extinction according to Rayleigh's scattering law is responsible for the reddish color of this illumination. During an eclipse, the lunar disk thus acts as a screen on which the distribution of illumination in the umbral and penumbral cones of the earth can be observed and measured. This is of geophysical interest as a method for the optical investigation of the earth's atmosphere at various altitudes, and astronomical observations of lunar eclipses are regularly made at many observatories. Such observations accomplish the following goals.

1. General description of the phenomenon. Here, the apparent brightness, color, and visibility of lunar surface details in the umbra and penumbra are specified. The results obtained by these highly primitive methods of observation are not easily amenable to astrophysical interpretation, particularly since they are strongly influenced by various subjective factors and by atmospheric and other conditions. Nevertheless, such observations show that the conditions for illumination of the lunar disk at equal distances from the cone axis differ for different eclipses: some eclipses are anomalously light, while others are anomalously dark.

2. Determination of the angular diameter of the earth's umbra. This is primarily done by observing the contact times, that is, the times of outer and inner contact of the lunar limb with the edge of the umbra. More extensive data are obtained from observations of the passage of the edge of the umbra across conspicuous lunar surface features, such as individual mountain peaks, central peaks in craters, etc. Blurring

of the umbra edge keeps the accuracy of these observations low, with errors of about 10 or 20 seconds. Observations of this type show that the semidiameter of the apparent umbra is on the average 2% larger than the geometrical shadow of the earth, due to the shadowing effect of the atmosphere.

3. Integrated-light photometry of the moon. Such measurements are made during almost every eclipse. The results are generally expressed as the difference m of the stellar magnitudes of the moon for a give phase of the eclipse and for full moon; the data are then plotted as a function of the phase or of the angular distance between the centers of the lunar disk and the umbra. Since the magnitude of the moon at true full moon can be obtained only by extrapolation to 0° phase angle, the results are often presented in terms of the difference m' between the magnitude of the moon during the eclipse and its magnitude before or after the onset of the penumbral eclipse. Such data make possible an objective comparison between the mean conditions of illumination in the umbra and penumbra for various eclipses.

4. Brightness measurements for individual areas of the lunar disk. These are made by photometric methods during an eclipse. If a single feature or a group of features (all with the same reflectivity) is measured during the whole eclipse, then we obtain a curve expressing the variation of illumination across a particular section of the umbra and penumbra. The results are represented in terms of the difference M of the brightnesses of the particular feature during the eclipse and at true full moon, or alternatively as the difference M' of the brightnesses during the eclipse and under extra-eclipse conditions. Quite a few observations of this kind have been made, but since they refer to different photometric systems, while the umbra has a distinct color, therefore a comparison of their results is quite difficult.

In order to interpret the results of photometric observations of lunar eclipses, the illumination-attenuation curve for M in the umbral and penumbral cone is compared with the results of theoretical calculations for various atmospheric models. The carrying out of such calculations constitutes a very complicated problem in atmospheric dioptrics, which cannot be solved without various assumptions concerning the vertical variation of both the refractive index of air and the extinction coefficient. Such theories have been developed by von Seeliger and Pagtzold in Germany, Richardson in England, Fesenkov in the USSR, and recently by Link and his co-workers in Czechoslovakia.

§86. Mars. The size and mass of Mars put it into the class of small planets. Due to the good visibility of its permanent surface features, the rotation of Mars has been studied more accurately than that of any other planet. For example, the sidereal rotation period, for which Ashbrook gives a value of $24^h37^m22^s.6689$ on the basis of observational material for the years 1704 — 1952, is known to at least within $\pm 0^s.003$. In ephemeris calculations, the north-pole coordinates are taken as

$$\alpha = 21^h11^m, \quad \delta = +54°39' \,(1950.0),$$

which corresponds to a 25° inclination of the equator to the orbit. The latest data, obtained by Camichel after processing many years of

photographic material collected at Pic du Midi Observatory, give

$$\alpha = 316^\circ.81, \quad \delta = +52^\circ.94 \quad (1950.0).$$

These values correspond to an equatorial inclination of 24°.80. Mars is thus very much like the earth as far as the length of the day and the succession of seasons are concerned, but a Martian year is 669 Martian days long.

Integrated-light photometry gives the magnitude of the planet as $-1^m.85$ (visual Harvard system) at mean opposition, which corresponds to a geometric albedo of 0.139. The phase coefficient is 0.015 stellar magnitudes per degree; the phase curve can be traced up to a phase angle of $\psi = 52°$, so that we can extrapolate fairly reliably to obtain the phase integral and the spherical albedo ($A_s = 0.154$).

The data on the color index of Mars are highly contradictory, and the values obtained by various authors range from $+1^m.3$ to $+2^m.1$. It has been suggested that these differences are partly due to actual fluctuations in the color of the planet, caused by variations in the atmosphere.

Martian topography has been studied in great detail. Three differently colored area types may be distinguished on the surface of Mars: 1) continenta (continents), all having medium reflectivity and a definite yellow-orange color; 2) maria (seas), which are less bright and have a different, less-saturated, color; the arms of these maria are called sini (bays) or freta (straits or channels); isolated spots the same color as maria are called laci (lakes), and a very small spot is called an oasis; dark rectilinear bands are named canals; 3) the polar caps or polar spots, finally, are very light, whitish patches surrounding the poles of the planet.

The outlines of the continents and maria are more or less permanent. A special branch of Martian research, known as areography, deals with the determination of the areographic coordinates (latitudes and longitudes) of these permanent features and with their mapping. Numerous catalogs of areographic coordinates have been published, as well as maps of the surface of Mars. Here, the different features have been named appropriately.

The continents have colors which do not change. The colors of the maria, on the other hand, vary, and these changes are clearly seasonal in nature. The color generally darkens in spring in each respective hemisphere. Subjective visual observations have repeatedly shown that the color changes from yellowish-brown to greenish or bluish. Objective measurements, however, show that the surface reflectivity of the maria always increases from the violet to the red, although less steeply than for continents. Therefore, the visually observed greenish or bluish colors are just an optical illusion due to the strong color contrast between the saturated orange of the continents and the brownish, though much less saturated, colors of the maria.

The white polar caps undergo even abrupter variations. As the vernal equinox approaches for a given hemisphere, the white cap begins to deteriorate and its boundary recedes toward the pole. By the time of the autumnal equinox, the white area will have shrunk to a minimum, after which it again begins to increase. Considerable data have been accumulated on the deterioration of the polar caps during different Martian years, on

the distinctive features of the process in the northern and southern hemispheres, and on the local characteristics of the boundary migration at different surface points on Mars.

Photographs of Mars taken through different light filters are very interesting. In 1909 Tikhov showed that, as we pass to shorter wavelengths, the brightness difference between the maria and the continents diminishes, while the contrast between the polar caps and the rest of the background increases; this is due partly to the veiling influence of the atmosphere and partly to the color properties of the surface. It has also been established that in violet and ultraviolet light only the polar caps are visible, while the other features vanish, so that the disk of Mars has a uniform appearance reminiscent of Venus. In blue light, the maria and the continents sometimes are not visible and sometimes are faintly discernible, which should be attributed to fluctuations of the atmospheric transparency.

The various properties described in the preceding have been the subjects of a thorough photometric analysis. As indicated in §80, the brightness B at any point on the surface of Mars is made up of two components: the light reflected from the surface and the light scattered in the atmosphere. At the time of true opposition, when the Mars-sun and Mars-earth directions practically coincide, so that the angle of incidence i equals the angle of reflection ε, the brightness factor ϱ (see §78) may be expressed as

$$\varrho = \frac{\pi B}{E} = r_0 \cos^q i e^{-2\tau \sec i} + \varkappa(\tau, i, r), \tag{4}$$

where r_0 is the brightness coefficient of the surface for a beam of rays arriving and reflected along the normal ($i = \varepsilon = 0°$), τ is the optical thickness of the atmosphere, and q is the smoothness factor, a special parameter proposed by Sytinskaya. The smoothness factor represents approximately the law of light reflection for the surface (for an orthotropic surface $q = 1$; for a pitted surface $q < 1$, possibly dropping to zero; for a specular surface $q > 1$). The function $\varkappa(\tau, i, r)$ includes all the brightness components connected with atmospheric scattering.

It is useful to observe a certain surface area on Mars for different values of the angle i, which constantly varies due to rotation. Then by assuming that the parameters in (4) remain constant in time, or alternatively by measuring the radial brightness distribution across the disk and assuming the conditions to be the same at all points measured, we obtain a system of equations of the form of (4). This system can then be solved for the parameters r, q and τ characterizing the surface and atmosphere of Mars in a certain spectral region. Such investigations were carried out by Barabashov in Khar'kov (who calculated the function $\varkappa$ using Schönberg's formulas), by Fesenkov (who proposed for $\varkappa$ a formula which took into account the illumination of the atmosphere by light reflected from the surface, but neglected higher-order scattering), and by Sytinskaya (who applied Sobolev's formulas, which take into account approximately both the luminous exchange between surface and atmosphere and higher-order scattering). As an example, Table 1 gives the results obtained using the 1939 observations. The good fit of the various sets of data for the surface, obtained using different means of calculating $\varkappa$, is a consequence of the low value of τ (a transparent atmosphere does not influence the surface brightness much).

TABLE 1

The true albedo r_0 for the surface of Mars and the atmospheric optical thickness τ (to a base 10)

Calculated for different models, according to 1939 observational data

Scattering indicatrix	Rayleigh				Rayleigh				Elongated				
Higher-order scattering	neglected				allowed for				allowed for				
Spectral region, m μ	r_0			τ	r_0			τ	r_0			τ	
	continents	maria	polar caps		continents	maria	polar caps		continents	maria	polar caps	continents	maria
380	0.057	0.057	0.30	0.205	0.051	0.051	0.32	0.145	0.064	0.064	0.34	0.117	0.117
430	0.092	0.083	0.28	0.175	0.039	0.089	0.30	0.101	0.103	0.092	0.31	0.074	0.071
560	0.190	0.163	0.28	0.103	0.134	0.159	0.28	0.047	0.179	0.159	0.29	0.034	0.034
630	0.238	0.160	0.24	0.080	0.236	0.160	0.24	0.031	0.235	0.161	0.24	0.028	0.025
730	0.283	0.155	0.18	0.051	0.278	0.159	0.18	0.023	0.277	0.159	0.18	0.018	0.018
q	1.22	1.20	—	—	1.18	1.04	—	—	1.16	1.08	—	—	—

Characteristically, the variation of τ with λ lags behind the curve prescribed by Rayleigh's law, which corresponds to a pure gas. Hence it follows that the atmosphere of Mars is in fact an aerosol, that is, it contains a substantial concentration of comparatively coarse suspended particles which are characterized by neutral scattering; this conclusion is in full agreement with all the other available data.

If the mass of the gas and the atmospheric pressure on a planet are proportional to the optical thickness τ (this is obviously nothing more than a working hypothesis), then by comparing the values of τ for the earth and Mars we can find the atmospheric pressure at surface level on Mars. Sytinskaya obtained in this way a pressure of 46 mm; Barabashov's data gave 37 mm; de Vaucouleurs, from his naked-eye estimates of the brightness, obtained 70 mm; Dollfus used polarimetric data to find 60 mm. The agreement between the different figures is better than would be expected. On the average, we may assume a pressure of 50 mm for aneroid-barometer readings (this being the actual pressure taking into consideration the weak gravitational acceleration). A mercury barometer would give a reading of 130 mm, from which it follows that the gas column per unit area on Mars is approximately 1/6 as massive as that on the earth.

It follows from the preceding that the atmospheric conditions at the surface of Mars correspond approximately to those prevailing in the earth's atmosphere at altitudes of 15 to 20 km, that is, they are stratospheric in nature. It must be kept in mind, however, that due to the low gravity the vertical pressure gradient β (formula (58), Chapter XI) will also be small, so that at a height of about 30 km the atmospheric pressures on Mars and on the earth will be the same, whereas at greater altitudes the pressure in the Martian atmosphere will actually exceed that for the earth.

Spectroscopic investigations of the chemical composition of the Martian atmosphere present a difficult problem. Numerous attempts to detect oxygen and water vapor, using Doppler-effect measurements to distinguish between lines of planetary and telluric origin, gave negative results.

Therefore, it follows that the upper limit for the content of these gases in the Martian atmosphere, relative to their content in the earth's atmosphere, is less than 0.2% for O_2 and less than 1% for H_2O.

So far, only carbon-dioxide gas has been definitely detected in the Martian atmosphere; its absorption bands were identified by Kuiper in the wavelength range between 1.5 and 2μ. A calibration made for the 1.605μ band by means of comparison with the CO_2 absorption spectrum in the earth's atmosphere gives an equivalent CO_2 layer 440 cm thick for Mars (pressure 760 mm, temperature 0°C), which corresponds to a partial pressure of 0.26 mm Hg.

It is thus clear that, although the amount of CO_2 in the Martian atmosphere is twice that for the earth, this gas is only present as a slight admixture to some other gases, apparently nitrogen or a nitrogen-argon mixture. Consequently, Mars, like Venus, has an atmosphere composed of nitrogen and carbon dioxide.

The cloud formations observed in the Martian atmosphere fall into two classes: 1) yellow clouds, similar to the continents in color, or else somewhat whitish or grayish; these are visible to the eye and show up on photographs taken in the yellow-red part of the spectrum; 2) violet turbidities, which show up only on ultraviolet photographs. Movements of the yellow clouds make it possible to study air currents in the Martian atmosphere, but the number of such observations made so far are too few.

A subject of lively discussion is the nature of, and the very existence of, a special "violet layer" in the Martian atmosphere. During the 1924 opposition, Wright (at the Lick Observatory) measured the diameter of Mars using photographs taken in various spectral regions; he found that the diameter increased toward the ultraviolet. Wright attributed this phenomenon to the fact that red-light photographs give the diameter of the planet's solid globe, while photographs taken in ultraviolet light give the diameter of the atmospheric envelope. This interpretation gave a thickness of up to 100 km for the Martian atmosphere. Later, Fesenkov showed that a gaseous atmosphere maintained in aerostatic equilibrium (§79) and having in it Rayleigh light scattering cannot produce this effect, since, if optically thick, it will be white-colored and not blue. In order to explain the Wright effect, we must assume that at a high altitude in the Martian atmosphere there is a thin layer which scatters the short wavelengths of the solar spectrum strongly. This "violet layer" has been associated with the overall luminous veil obscuring the planetary surface on ultraviolet photographs and also with the violet clouds. The latter are considered to be local condensations of the violet layer. Some very different assumptions have been advanced to explain the constitution of this layer; it has been suggested that the violet layer is composed of minute H_2O or CO_2 crystals, or meteoric dust, etc. Some authors endow it with extraordinary optical properties, such as an increase of scattering power in the ultraviolet in excess of that prescribed by Rayleigh's scattering law, or, conversely, weak scattering and strong absorption of short wavelengths.

An alternative approach claims that the Wright effect is not a real phenomenon, that it is produced by purely photographic effects related to the fact that in red light Mars shows a very sharp limb darkening which does not appear in the ultraviolet. Barabashov attributes the observed

difference in the planet's diameters to an underexposure of the limb areas in red light, while Sharonov has blamed it on photographic irradiation, which increases with the limb brightness.

The temperature conditions on the surface of Mars have been studied theoretically by Milankovič and Hess, and experimentally by various American observers. The results agree on the whole and they may be summed up briefly as follows. The mean disk temperature at opposition is close to −30°C. At the subsolar point (the disk center) it rises to +16° or +20°C, the dark maria being from 5 to 10° warmer than the continents. At the evening terminator the temperature drops to 0°C even at the equator, and at the morning terminator it is as low as −40° or −50°C. In the winter hemisphere the daytime temperature ranges from −50° to −80°C, and in the zone of uninterrupted summer day it is from 0° to +10°C. The latter value is of great significence for interpretation of the white polar patches.

The various conclusions concerning the nature of the Martian surface are based on simple eye estimates of the reflectivity (brightness, and color of surface features during visual telescopic observations), or on determinations of the brightness coefficients in various spectral regions.

The continents are generally regarded as desert areas. Their smoothness factors q are close to unity, indicating an exceptionally smooth surface. The common opinion is that their surface is made up of sand, like the sandy deserts on the earth. However, a colorimetric comparison made by Sharonov has shown that the sands of terrestrial deserts are colored much less strongly than the surface of Mars (for sands the mean yellowness index is $D = 0.49$, and for Martian continents it is $D = 1.09$); only ocher limonites (the earthy, crumbly variety of the mineral limonite, $Fe_2O_3 \cdot nH_2O$, for which $D = 1.10$) are close in this respect to the surface of Mars. Dollfus, in his polarimetric comparison of the Martian surface with terrestrial materials, also found no similarity to sands and clays, but he found a close correspondence with the properties of powdered limonite.

These conclusions are consistent with the extraordinary phenomena observed on Mars in 1956, when persistent orange-colored mists obscured considerable parts of the disk, including the polar cap, for a few days in succession. Sandstorms cannot persist for such long periods, but fine limonite dust can remain suspended in the atmosphere for a considerable time.

Researchers vary in their opinions concerning the nature of the maria. The main difficulty is how to explain the seasonal variations in reflectivity. The original hypothesis of Schiaparelli, according to which the dark areas on Mars are actually water bodies, has been rejected, since a water surface would be observed as a star-shaped reflection of the sun. The moistening hypothesis, according to which the surface cover becomes darker because it absorbs moisture, is supported by some authors, but it is definitely inconsistent with the absence of water vapor in the Martian atmosphere. McLaughlin's volcanic hypothesis is currently being discussed in scientific circles. According to this theory, the dark cover of the maria is volcanic ash, and the seasonal deposition and removal of the ashes is connected with the seasonal variation of the atmospheric air currents. The most popular point of view is the hypothesis explaining the color changes as being due to the growth and withering of vegetation. Most authors regard this vegetation

as a type of lower plant similar to lichens and algae, but Tikhov and his co-workers maintain that higher forms of vegetation may also exist on Mars.

Sinton's recent investigations are of considerable interest in connection with the problem of Martian vegetation. He found that in the far infrared the maria have wide absorption bands centered at 3.43, 3.56, and 3.67μ; these correspond to the absorption bands of organic molecules. However, it has still not been proven that a mineral cover could not produce these bands.

The question of the Martian "canals" remains open; some observers reject out of hand the possibility of their existence, interpreting the phenomenon as an optical illusion, while others maintain that these are strictly geometrical formations. Most researchers, however, are of the opinion that although these long rectilinear formations actually exist on Mars, they are of natural origin and have no special regularity.

Radio-astronomical observations on 3.15-cm waves have detected a weak radio emission of constant intensity coming from Mars. It is apparently purely thermal in origin, and it corresponds to a mean ground temperature of −50°C.

Seasonal effects must also be taken into consideration in explaining the existence of the polar caps. Of the four main hypotheses, namely snow and ice cover, cloud-fog cover, solid carbon dioxide, and white-salt efflorescences, only the first two are now considered seriously. The CO_2-snow hypothesis has been rejected due to the unsuitable temperature conditions (CO_2 solidifies only at −78°C, which is apparently attained on Mars in winter months only) and also because according to Kuiper's data the polar caps have a low reflectivity in the far infrared; this could describe frozen H_2O but does not correspond to solid CO_2, whose albedo is high in the long-wave region also.

§87. **Jupiter.** As the planet with the largest mass and the largest dimensions, Jupiter is of special interest. Taking the corresponding values for the earth as unity, we have the following values for Jupiter: mean radius, 10.95; volume, 1312; and mass, 316.94. A comparison of the last two quantities gives a mean density of 1.34 g/cm^3, or 0.24 of the mean density of the earth. The acceleration of gravity on Jupiter is only 2.64 times that on the earth, but the escape velocity is high, equal to 60 km/sec.

Since it is flattened considerably, the disk of Jupiter appears as an ellipse in the telescopic field of vision. Direct measurements indicate a flattening of 1/16, while G. Struve's calculations based on an analysis of satellite perturbations give a value of 1/15.4

Jupiter is the first, and most characteristic, representative of the group of four "Jovian" planets, which also includes Saturn, Uranus, and Neptune. These planets differ from the previously considered "terrestrial" planets in the following ways: 1) they have larger masses and radii, but lower densities (close to unity); 2) the axial rotation is more rapid; 3) there is considerable flattening; and 4) they have low values of the parameter $\frac{e}{q}$ (formula (29), Chapter XI), indicating high concentrations of matter at the center and low densities in the peripheral layers. The structures of the Jovian planets are thus radically different from the structures of the terrestrial planets. Recently, some authors (including Fesenkov, Kozyrev, and Masevich) published structural models for the Jovian planets, based on the theory of stellar interiors. They assumed that the main bulk of the

planetary mass consists of hydrogen, which is maintained at very high pressures in the central parts.

Integrated-light photometry indicates a visual magnitude of $-2^m.29$ at mean opposition; the corresponding geometric albedo is $A_g = 0.375$ and the illustrative albedo is $A_i = 0.56$. The phase angle of Jupiter does not exceed 12°, so that the phase curve is not known; the phase coefficient, however, has been found to be 0.015 per degree. The color of Jupiter remained uncertain for a long time; Russell's summary of data gives a color index of $+0^m.50$ which would indicate a color more bluish than the sun. The more recent visual-colorimetric observations of Radlova and others, however, have established definitely that Jupiter is yellower than the sun by $0^m.1$ to $0^m.2$ units of color index.

The large disk of Jupiter has many different features on it. The most pronounced of these are the dark and light bands, whose boundaries follow the planetographic parallels. According to accepted terminology, the dark-colored bands are called belts and the bright gaps zones. Against the background of belts and zones may be observed various rounded bright spots, isolated dark formations shaped like oval spots, and various elongated strips which intersect the belts and zones at different angles.

The rotation of Jupiter is zonal: each band has its own rotation period, the length of the periods increasing with the planetographic latitude. For the equatorial zone, the rotation period is approximately 9^h50^m, while for belts at latitudes of 35 to 40° it increases to 9^h56^m. Repeated attempts to establish a more accurate law describing the latitudinal variation of rotation velocity did not give any definite results, since different features in the same zone have different rotation periods.

The zonal nature of the rotation complicates the reckoning of longitudes on Jupiter. Two systems of longitudes are used. One of these, called the first system, corresponds to a period of $9^h50^m30^s.003$, corresponding on the average to the rotation of features in the equatorial zone. The other system, known as the second system, has a period of $9^h55^m40^s.632$, derived from observations of the so-called Red Spot.

The equatorial elements of Jupiter have been reliably determined. The ephemerides are calculated using the following coordinates of the planet's north pole, which were obtained by G. Struve at Pulkovo Observatory from the theory of satellite motion and which are also in good agreement with visual observations of the migration of spots across the disk:

$$\alpha = 17^h52^m0^s.84, \quad \delta = +64°33'34''.6 \quad (1910.0).$$

This corresponds to a 3° inclination of the equator to the orbit, so that there are practically no seasons on Jupiter.

The nature and the positions of features on Jupiter's disk constantly change, sometimes quite rapidly, which is consistent with the accepted opinion that the surface of the planet is an atmospheric cloudy layer. The planet must therefore be kept under continuous surveillance. At present such a "Jupiter Survey" is being carried on by amateur observers (for example, the Jupiter Section of the British Astronomical Association and the VAGO* Observers' Collective). The results are published in the form of synoptic charts of the planet's surface for each opposition, or for shorter

* [Vsesoyuznoe astronomo-geodezicheskoe obshchestvo (The All-Union Astronomical and Geodetic Society)]

time intervals, and also in the form of tables listing the latitudes of the boundaries of the belts and zones, and the coordinates and rotation periods for individual smaller features.

The extensive observational material on hand at present indicates that the overall variability of the planet's disk has a certain irregular periodicity or cyclicity. Epochs during which the planet's equatorial region is occupied by a bright zone with wide dark belts lying north and south of it give way to periods during which the two dark belts merge into a single dark-colored equatorial belt; this process is repeated at intervals of 9 to 11 years. Since the length of this cycle is close, on the one hand, to the sidereal revolution period of Jupiter (11.9 years) and, on the other hand, to the 11-year cycle of solar activity, many investigations have been made attempting to establish a correspondence between the different cycles, but no definite results have been obtained. It has also been observed that the alteration in the distribution of zones and belts is accompanied by an overall enhancement or weakening of the variation activity observed on Jupiter; in addition, it affects the planet's stellar magnitude and color index.

Besides transient formations, Jupiter also has some semipermanent surface features. The most famous of these is the so-called Red Spot, discovered in 1878. Although it varies in brightness and color, the Red Spot has nevertheless been observed continuously since that date. It is shaped like an oval, elongated along the parallel, and it is 50,000 km long and 10,000 km wide. The Red Spot remains at approximately the same latitude (about $-20°$ for the center), but it goes through complex changes in rotation period, so that its longitude in the second system varies constantly and irregularly. It is noteworthy that another comparatively stable formation exists at the same latitude, namely the "southern veil" or "south tropical disturbance." This is a dark region, with a rotation period somewhat shorter than that of the Red Spot. Consequently, the "tropical disturbance" overtakes the Red Spot once every two years; then, it flows around the spot in two streamers branching to the north and to the south, an event which is accompanied by a number of peculiar phenomena.

The disk of Jupiter has been the subject of numerous photometric and colorimetric investigations, including visual (Schönberg, Fesenkov), photographic (Barabashov and co-workers, Sharonov), and photoelectric (Maggini) studies. The results indicate that the radial distribution of brightness across the disk (mostly over the equator) follows an approximate cosine law, so that the surface is close to orthotropic in its reflective properties. The color varies away from the center (turning bluer); at equal distances from the center the color is redder for features of low brightness.

A characteristic feature of Jupiter's spectrum is the presence of distinct absorption bands, giving a definite indication of the type of absorption in the atmosphere of this planet. The strongest band lies in the red region between 6450 and 6507 Å; it has a complex structure. Although these bands were discovered at the end of the 19th century, the chemical nature of the parent gases remained unknown for a long time; it was only in 1931 that Wildt showed the strongest bands to be methane (CH_4) bands and some of the less intense bands to be ammonia (NH_3) bands. This conclusion was verified in detail at Mount Wilson Observatory by a painstaking comparison of the spectrum of Jupiter (as well as those of Saturn, Uranus, and Neptune) with the absorption spectra of gases, observed under laboratory conditions

through a 70-foot tube. For NH_3, 30 lines were identified near 6450 Å and 39 lines near 7900 Å. A comparison of line intensities gave for Jupiter an equivalent homogeneous NH_3 layer (T = 0°C, pressure 760 mm Hg) about 5 to 10 m thick. An analagous comparison for methane gave an equivalent layer of 150 m.

Methane and ammonia are clearly not the main constituents of Jupiter's atmosphere. They account for only a small fraction of the atmospheric gases, but the other components are not observed spectroscopically. These may be nitrogen, argon, helium, and hydrogen. The hypothesis of hydrogen atmospheres for the Jovian planets is the most popular one. In 1952, it gained additional support from Kuiper's discovery that in the spectra of Uranus and Neptune there is a weak, diffuse 8279 Å band, which was identified with the H_2 band observed by Herzberg under laboratory conditions.

Information on the constitution of the Jovian atmosphere has also been obtained from observations of the occultation of stars by Jupiter. In particular, during an occulation of the star σ Arietis (20 November 1952), Baum and Code used the Mount Wilson 60-inch reflector to make a photoelectric recording of the attenuation of the stellar brightness. As observed in §80, the variation of stellar brightness under these conditions is determined by the vertical variation of refraction in the planetary atmosphere, and thus (since the refractive index is related to the gas density) by the density gradient β (chapter XI, formula (59)); the latter parameter can therefore be found from occultation observations. For the isothermal atmospheric model it was found that $\beta = 0.12$ per km. Furthermore, given the acceleration of gravity and assuming a probable temperature value, we can calculate the molecular weight μ. For T = 86°K, we obtain $\mu = 3.3$, which proves the high content of light molecules in Jupiter's atmosphere (6 H_2 molecules for each CH_4 molecule gives $\mu = 4$, while 12 H_2 molecules for each CH_4 molecule gives $\mu = 3$).

The remoteness of the planet from the sun is responsible for the low equilibrium temperature: $T_1 = -100$°C and $T_2 = 151$°C. These values are close to the radiometric temperature at the center of the disk, −140°C, found by Coblentz.

An interpretation of the various observational data mentioned above is considerably difficult. We have already indicated that the apparent surface of the planet is a layer of cloudlike aerosol, floating at a certain depth in the gaseous atmosphere. Since the freezing point of CH_4 is −184°, while that of NH_3 is −78°C, it is natural to assume that the disperse phase of the aerosol consists of ammonia crystals.

The limb darkening was previously explained as an effect of extinction in the gas layer above the clouds, the attenuation increasing toward the limb because the path of a ray in the gas layer becomes longer. However, this explanation applies if the gas attenuates the transmitted radiant flux by means of absorption only, and not by scattering, that is, if the brightness factor η of the gas is close to zero. Since this is unlikely for the visible spectrum, this explanation has been rejected, particularly since, according to Ambartsumyan's theory of scattering, a purely scattering medium (regardless of whether it is a gas or an aerosol) produces a limb darkening close to that actually observed for Jupiter and for other Jovian planets.

Intensity measurements for the CH_4 and NH_3 absorption bands at various distances from the disk center showed virtually no radial variation. Hess

has interpreted this result in the light of the previous opinions; he suggests that the height of the cloud layer increases, while the thickness of the gas layer above it correspondingly decreases, toward the edges of the disk, so as to compensate for the increase in path length. Since the scattering is also appreciable at the wavelengths of the absorption bands, it is more reasonable to explain this fact from the point of view of scattering theory, assuming that the limb darkening due to scattering is the same inside the absorption bands and outside of them.

It is difficult to explain the peculiar coloring of the surface features of Jupiter. Wildt has suggested that the aerosols contain special coloring compounds, while Sharonov maintains that the different colors may be the result just of appropriate combinations of Rayleigh and neutral scattering.

The decameter radio waves emitted by Jupiter were studied starting in 1950. However, it was only the series of special radio-astronomical observations carried out by Burke and Franklin in 1955 which established definitely that Jupiter is a very powerful radio source. Jupiter emits radio bursts which last from 0.5 to 1 sec, and these appear and disappear periodically, varying with the longitude of the central meridian. The period was found to be close to the rotation period of the Red Spot; the summary of observations prepared by Douglas gives a mean period for this of $9^h55^m29^s,37$. A third system of Jovian longitudes has thus been introduced to agree with this radio period. Here, the initial meridian is defined as the one which coincided with the initial meridian of the second system at 0^h on 1 January 1957.

The intensity curve for the decameter radio waves shows that there are probably from two to four independent radio sources. If these were orthotropic, then the waves from each source would be detected in a 180° interval of central-meridian longitudes. Actually, the reception range is limited to 50 or 60°, and it decreases at higher frequencies. The radio beam is thus fairly narrow. This can be attributed to the presence of a quite powerful ionosphere: the waves emitted by a source below the ionosphere experience total internal reflection at the lower boundary of the ionospheric layer for certain zenith distances, and they do not emerge from the atmosphere. This total reflection also determines the angular span of the reception cone. The radio waves received are circularly polarized, apparently due to certain ionospheric processes.

Besides radio bursts, there are three other types of radio emission from Jupiter: thermal emission, corresponding to a constant radio-brightness temperature of −140°C; nonthermal emission which is unpolarized; and nonthermal emission which is polarized. The nonthermal radio waves, both polarized and unpolarized, are very useful for analyzing the physical processes which probably occur in Jupiter's ionosphere.

§88. Saturn and its rings. The figures in Table 1 of Chapter XI show that this planet is the second largest in the solar system; its characteristic measurements are as follows (earth = 1): equatorial diameter 9.02, volume 734 and mass 94.4. Saturn is unique with regard to some of its other characteristics. It has the lowest mean density in the solar system, 0.71 g/cm^3 (or 0.13 of the earth's density), and the highest oblateness $^1/_{10}$. Finally, Saturn has the lowest value of the parameter $\frac{e}{q}$, namely 0.62, which is close

to the theoretical lower limit of 0.5 and which indicates an exceedingly high concentration of mass at the center.

Integrated-light photometry and colorimetry studies of Saturn are complicated by the presence of the rings. The ring brightness varies to a great extent with the angle of elevation D of the earth above the ring plane, and also partly with the phase, although the phase angle ψ never exceeds 6°. The visual magnitude m of Saturn at its mean distance (Δ = 9.5388 A.U., L = 8.5388 A.U.) is generally expressed by the formula

$$m = +0.79 - 2.5965 \sin D + 1.2526 \sin^2 D + 0.0436 \psi. \qquad (5)$$

The data for D = 0 and ψ = 0 give albedo values of A_g = 0.42 and A_i = 0.63, which indicates a high reflectivity due to the cloudy nature of the surface. Saturn is on the average yellow in color, its color index being variously quoted as $+1^m.12$ (Russell), $+1^m.22$ (King), and $+1^m.18$ (Radlova), which is approximately $0^m.3$ to $0^m.4$ higher than the color index of the sun.

The planet's disk has very few visible surface features. The relatively bright equatorial zone has dark bands adjoining it. The polar regions of the planet also appear to be darker. Individual bright or dark spots are very seldom observed, which complicates studies of the planetary rotation. Observations of such spots have led to the conclusion that Saturn, like Jupiter, has a zonal type of rotation; the equatorial period is 10^h14^m, the period in the latitude belt from +35° to 40° is 10^h38^m, and for latitudes of ±57° it is more than 11^h.

The color characteristics of the surface have been established using photographs taken in narrow spectral regions. In red and infrared light, no surface features are observed, but the limb darkening is enhanced; in yellow and green light, the equatorial belt is at its brightest; and in the short-wave region, the equatorial belt and the polar areas appear to be very dark while the limb darkening is diminished.

Several studies based on the surface photometry of Saturn's disk were conducted by Barabashov and his co-workers at the Khar'kov Observatory, and these provided a quantitative illustration of the color and brightness phenomena. The absolute surface photometry of the disk carried out by Sharonov at Pulkovo, using photographs taken with a 30" refractor, gave a brightness coefficient of 0.68 in the photovisual system for the center of the disk.

The spectrum of Saturn shows the same bands as the spectrum of Jupiter. However, the CH_4 bands are enhanced, corresponding to an equivalent gas layer 350 m thick, while the NH_3 bands are weaker, giving an equivalent layer thickness of only 2 m.

Saturn's rings are unique in the solar system. Since the rings are tilted at 28° to the plane of the ecliptic, their appearance as viewed by an earth-bound observer varies with the angle D between the ring plane and the line of sight. In the epoch when the full ring spread is visible, the rings have the form of a wide ellipse with an axial ratio of 1 : 0.88, and the northern and southern edges of the disk are projected onto the rings. The angle D then decreases, and after 8 years it goes to zero, that is, the earth passes through the ring plane. At this stage the rings show up as a thin line, which even disappears completely for some time, indicating that the ring thickness is very small. The rings then turn the other side toward the observer, and the full spread of the rings is observed again after 15 years, the cycle subsequently repeating itself.

A simple visual observation shows that Saturn's rings may definitely be divided into the following three zones: an outer ring *A*, separated by a sharp black gap (called Cassini's division) from the middle ring; the brighter middle ring *B*, separated by a dark gap from the inner ring; and the semitransparent inner ring *C*, known as the "crepe" ring. Rings *A* and *B* have several narrow divisions, less sharp gaps, and gradations, one of which is the Encke division in ring *A*.

The position of the ring plane is specified by the following coordinates of its north pole, derived by G. Struve at Pulkovo:

$$\alpha = 36°12'.0 \quad \delta = +83°2'.7 \quad (1899.0).$$

These coordinates are also assumed to specify the north pole of the planet's axis of rotation. The dimensions of the ring system are as follows:

Outer radius of ring A	275,000 km
Inner radius of ring A	258,000 "
Width of Cassini's division	5,000 "
Outer radius of ring B	253,000 "
Inner radius of ring B	227,000 "
Gap between B and C	1,600 "
Inner edge of ring C	109,000 "

The rings are exceptionally thin, no more than 15 to 20 km in width. This follows from the fact that, when the earth passes through the plane of the rings, they disappear for some time and are not visible even through the largest telescopes. Moreover, not only is the crepe ring *C* (through which the planet's globe shows up clearly) semitransparent, but also the two other zones are not entirely opaque either. This is obvious from observations of the occultations of stars by the rings, and also from the visibility of Saturn's satellites as they pass through the ring shadow.

The mass of the rings is negligible. According to G. Struve, the rings do not exert any perturbing influence on the motion of the satellites, so that their mass cannot be greater than $\frac{1}{4}$ of that of the moon.

The structure of the rings was analyzed theoretically by Maxwell and independently by Kovalevskaya. They proved that neither a solid nor liquid formation of this shape could be maintained in equilibrium and remain stable. Therefore, the only possibility is that the rings have a meteoritic structure, and that a swarm of small particles, each moving independently around Saturn in a circular orbit, is spread out to form the thin ring layer. This hypothesis was brilliantly confirmed by the spectroscopic investigations of Belopol'skii, Keeler, Campbell, and Deslandres, who determined the line-of-sight velocities and revolution periods for various ring zones and showed that the periods increase steadily with the distance from the center of Saturn, in full quantitive agreement with Kepler's third law.

The meteoritic theory explains the existence of empty gaps such as Cassini's division. Kirkwood has shown that particle motion at distances from Saturn where the particle period is commensurable with the revolution period of one of the satellites is necessarily unstable, due to the gravitational perturbation of the satellite. Actually, the distance of Cassini's division from the planet is such that the corresponding revolution period of particles in it would be $\frac{1}{2}$ the period of Mimas, the satellite closest to the ring, $\frac{1}{3}$ the period of Enceladus, and $\frac{1}{4}$ that of Tethys.

The sizes of the constituent particles of the rings may possibly be determined by means of photometric investigation. The important photometric properties of the rings are, first, constancy of brightness regardless of the actual angle of incidence of solar rays and, second, a rapid drop in brightness with increasing phase angle (the latter property is expressed by formula (5) for the integrated brightness of the planet). The phase coefficient for just the rings is 0.4 to 0.5, while the phase coefficient for Saturn itself is too small to be determined.

The first photometric theory of the rings which attempted to explain these properties was that of the German astronomer von Seeliger. He examined a model of a medium consisting of orthotropic spheres, uniformly distributed in a vacuum. At a phase angle of $\psi = 0°$, each element of the medium hides its own shadow, so that the brightness is maximum. Then, as ψ increases, the shadows emerge onto the image plane, and the shadowing effect rapidly lowers the apparent brightness.

However, this simple "shadow" theory is not adequate to explain more recent measurements of the ring brightness under various conditions of illumination and observation (Schönberg, Fesenkov). Schönberg found a certain similarity between the brightness variation of the rings and that calculated according to Mie's theory for dielectric scattering elements with radii of 3.6 μ.

A more exact theory must take into consideration, not only the effect of reciprocal shadowing of the particles, but also the luminous exchange between particles, that is, higher-order scattering. Bobrov has studied this problem, and his calculations indicate that the elements constituting rings *A* and *B* must be comparatively large, about the size of stones and small boulders,

§89. Uranus, Neptune, and Pluto. Visual observations of the disk of Uranus are difficult due to the small angular size and the low brightness. The observations which have been made have led to contradictory results: some observers report a system of equatorial bands, as on Jupiter, while others report rounded spots.

The main astrophysical data on Uranus pertain to its brightness variations. Older data disagree considerably; for example, for m_0 Müller obtained visually $5^m.86$, Pickering obtained $5^m.51$, and Russell assumed $5^m.74$, the latter value giving $g = 6.98$ and for $\delta_1 = 36''.0$ leading to an albedo of $A_g = 0.42$. Since neither the phase curve nor the phase coefficient are known (the phase angle of the planet does not exceed 3°), the albedo is calculated using $q = 1.5$, which gives $A_1 = 0.63$. Of the more recent determinations, we should mention the data of Parenago, who used a large series of visual observations to obtain the value $m_0 = 5.64$, and also the very accurate photoelectric measurements of Giclas in 1950 — 1952, which in the conditionally visual system gave $m_0 = 5.52$ and a color index of $0^m.54$.

Short-period brightness fluctuations have made it possible to determine the rotation period of Uranus, for which Parenago gives 10^h49^m. This value is also confirmed by spectroscopic observations, which indicate retrograde rotation. The position of the axis is specified by the following coordinates of its pole:

$$\alpha = 4^h21^m, \quad \delta = +29°.8,$$

which gives a 98° tilt of the equator to the orbital plane. Uranus is the only planet in the solar system which has a retrograde rotation and whose axis lies almost in the plane of its orbit.

The succession of seasons on Uranus is indicated by the integrated brightness. When the axis points toward the sun and the earth, the short-period brightness fluctuations cease, as was shown, for example, by the 1950—1952 photoelectric observations. Furthermore, a long-period brightness variation has been detected, the period being half the time for the revolution of the planet about the sun; brightness minima occurred in 1880 and 1921, when the rotation axis was perpendicular to the radius vector, and brightness maxima occurred in 1900 and 1942, when the axis pointed toward the sun. The amplitude of these fluctuations, according to Becker, is close to $0^m.3$. The brightness variations are due to the considerable flattening of the planet ($^1/_{18}$), on account of which the projection area of its disk onto the image plane will vary, and possibly due to color differences between the various latitude belts as well.

The absorption spectrum of Uranus's atmosphere contains methane bands which are stronger than those in the spectrum of Saturn; their position in the spectrum is responsible for the greenish color of the planet's disk as viewed with the eye. The intensity of the bands corresponds to a 1500-m layer of CH_4. No ammonia bands have been detected. The diffuse 8270 Å band found by Kuiper has been identified by Herzberg with the rotational band of molecular hydrogen; its intensity corresponds to an equivalent H_2 layer 18 km thick.

Radiometric observations of Uranus indicate a very low planetary heat, corresponding to a temperature lower than −200°C. The equilibrium temperatures of the planet are −184° and −200°C.

Neptune is very similar to Uranus in its geometrical, gravitational, and physical properties, although it has a somewhat larger mass. The diameter determinations have been contradictory; older data showed Neptune to be somewhat larger than Uranus, while newer measurements indicate the opposite (see Tables 1 and 2, Chapter XI). According to Russell, Neptune's diameter is 3.92 earth diameters and according to Rabe it is 3.89 earth diameters, the latter value also giving a somewhat higher mean density, $1.58\ g/cm^3$. An even smaller diameter, namely 3.50 earth diameters, and a proportionately higher density of $2.22\ g/cm^3$, were obtained by Kuiper in 1948 from observations made with a disk micrometer; however, this result requires further verification. A flattening of 1/50 was calculated using celestial-mechanical relations.

Repeated attempts to observe surface features on Neptune have given no definite results; therefore, if surface features do exist, then their contrasts are even lower than those on Uranus. Consequently, the Jupiter-Saturn-Uranus-Neptune sequence is characterized by a regular decrease in feature visibility with increasing distance from the sun.

Neptune's disk has not yet been studied by means of surface photometry, but many integrated-brightness measurements have been made, giving an apparent magnitude of $m_0 = +7^m.15$ (for Δ = 29.07 A.U. and L = 28.07 A.U., according to Russell's summary) and a color index close to the sun's, indicating approximately neutral scattering (although the disk, when observed visually, has a distinct greenish color). These figures correspond to an illustrative albedo of 0.52. If we use Schatzel's latest photoelectric

determinations, expressed in the international photovisual system of units, and if we take Kuiper's diameter, then we find that $m_0 = 7^m.95$ and $A_i = 0.49$.

Short-period brightness fluctuations attributable to axial rotation have been repeatedly observed for Neptune. All the observational material on this subject was recently reviewed by Günther, who obtained a period of $0^d.53020 = 12^h43^m$ and $m_0 = 7.85 \pm 0.01$. The brightness curve is quite complicated and varies with time. It cannot be explained by the rotation of a body with simple surface spots, and Günther has therefore suggested a complex hypothetical model with a two-layer atmosphere and windows in the upper layer, through which the specularly reflecting lower layer is visible.

The photometric rotation period of Neptune differs somewhat from the $15^h.8$ period found by Menzel and Moore on the basis of spectroscopic measurements, which showed that the planet has direct rotation. The latter fact is of considerable cosmogonic significance, since the larger satellite, Triton, moves in a retrograde motion.

Neptune's spectrum is characterized by very narrow CH_4 bands, which are even stronger than those in the spectrum of Uranus, corresponding to an equivalent methane layer 2500 m thick. Ammonia bands are absent, but the 8270 Å hydrogen band is observed. The spectra of the four Jovian planets thus fall into a sequence, in which the intensity of the CH_4 bands increases with the distance from the sun, while that of the NH_3 bands drops off. This can be explained by a hypothesis according to which the NH_3, crystals of which apparently form the cloud cover of Jupiter (and possibly of Saturn as well), freezes as the temperature drops and then precipitates to the surface, exposing a deeper-layer lying of gaseous CH_4.

The only reliable physical characteristics of Pluto are those obtained by integrated-light photometry. The visual magnitude m_0 was found to be $14^m.71$ (Graff), the visual-photoelectric magnitude is $14^m.90$ (Walker and Hardy), and the photovisual and photographic values are $14^m.74$ and $15^m.41$ (Baade), which corresponds to a color index of $+0^m.67$ (a neutrally colored surface).

The photoelectric observations made by Walker and Hardy in 1953—1955 indicated periodic fluctuations of Pluto's brightness with an amplitude of over $0^m.1$ and a period of $6^d.390$; the latter figure is accepted by the authors as the rotation period of Pluto.

The mass of Pluto has been determined from the perturbations in the motion of Neptune; it was found to be 0.99 ± 0.09 of the earth's mass. The question of Pluto's size, however, has not been settled yet. Kuiper maintains that he successfully resolved and measured Pluto's disk, obtaining a diameter of about 0.5 of that of the earth. This gives an albedo of 0.17, which is a plausible value; however, when coupled with the preceding value of the mass, this gives a mean density of about 45 g/cm^3. If we assume that Kuiper's diameter is too low, perhaps due to strong limb darkening, then we obtain an exceedingly low albedo (if Pluto's diameter is equal to the earth's, then the albedo is 0.04). A study of all the data for Pluto led Caillatte to suggest the following elements as a basis for discussion:

mass, 0.99 of the earth's;
radius, 7200 km;
density, 3.9 g/cm^3;
geometric albedo, 0.05.

The solar irradiation at the mean distance of Pluto from the sun is 1/1600 of that at the earth's orbit, which corresponds to equilibrium temperatures of $T_1 = -211°C$ and $T_2 = -229°C$.

§90. Planetary satellites. Thirty-one satellites are known at present in the solar system. They are listed, together with some of their orbital elements, in Table 2.

TABLE 2

Planets and satellites	Discoverer and year of discovery	Mean distance from planet (1000 km)	Sidereal period of revolution (mean days)	Inclination of orbit to planetary equator	Eccentricity of orbit	m_0	g	Diameter km
EARTH								
Moon	—	384	27.32	18°—29°	0.055	−12.55	+0.46	3476*
MARS								
Phobos	Hall, 1877	9.4	0.32	1°	0.017	+11.5	+9.1	15
Deimos	Hall, 1877	23.6	1.26	2	0.003	13.0	+10.6	8
JUPITER								
V Amalthea	Barnard, 1892	181	0.50	0	0.003	14.0	+7.3	160
I Io	Galileo, 1610	422	1.77	0	0.000	5.54	−1.16	3730*
II Europa	Galileo, 1610	671	3.55	0	0.000	5.69	−1.01	3150*
III Ganymede	Galileo, 1610	1,070	7.15	0	0.001	5.08	−1.62	5150*
IV Callisto	Galileo, 1610	1,881	16.69	0	0.007	6.28	−0.44	5158*
VI	Perrine, 1904	11,400	250.62	31	0.155	14.7	+8.0	130
VII	Perrine, 1905	11,700	260.07	30	0.207	17.5	+10.8	40
X	Nicholson, 1938	11,700	254.2	—	0.140	18.8	+12.1	24
XII	Nicholson, 1951	21,000	620	—	0.130	19.0	−12.3	22
XI	Nicholson, 1938	22,600	692.5	—	0.207	18.5	+11.8	30
VIII	Mellote, 1908	23,500	738.9	145	0.378	16.9	+9.8	56
IX	Nicholson, 1914	25,100	745.0	154	0.275	18.0	+11.3	27
SATURN								
I Mimas	Herschel, 1789	185	0.94	2	0.019	12.1	+2.5	650
II Enceladus	Herschel, 1789	238	1.37	0	0.005	11.6	+2.0	800
III Tethys	Cassini, 1684	294	1.89	1	0.000	10.5	+0.9	1300
IV Dione	Cassini, 1684	377	2.74	0	0.002	10.7	+1.1	1200
V Rhea	Cassini, 1672	527	4.52	0	0.001	10.0	+0.4	1750
VI Titan	Huygens, 1655	1,200	15.95	0	0.029	8.3	−1.3	4200*
VII Hyperion	Bond, 1848	1,500	21.28	1	0.119	13.0	+3.4	500
VIII Iapetus	Cassini, 1671	3,600	79.33	14	0.028	10.1	+0.5	1800
IX Phoebe	Pickering, 1898	12,900	550.45	149	0.166	14.0	+4.4	250
URANUS								
V Miranda	Kuiper, 1948	130	1.41	0	—	17.0	+4.9	—
I Ariel	Lassell, 1851	192	2.52	0	0.007	15.5	+3.4	800
II Umbriel	Lassell, 1851	267	4.14	0	0.008	15.8	+3.7	640
III Titania	Herschel, 1787	488	8.71	0	0.002	14.0	+1.9	1600
IV Oberon	Herschel, 1787	586	13.46	0	0.001	14.2	+2.1	1450
NEPTUNE								
I Triton	Lassell, 1846	354	5.88		0.000	13.6	−1.2	4800
II Nereid	Kuiper, 1949	8,000	359.4		0.76	19.5	+4.7	320

* The satellite diameters which have been measured micrometrically are marked with an asterisk; the rest of the values are hypothetical.

The nature of the moon was discussed in Chapter X. The other 30 satellites are celestial bodies with exceedingly small apparent angular diameters, and they must therefore be studied using the same methods as are used to study the asteroids.

Through a large telescope, only the four largest satellites of Jupiter (discovered by Galileo and called the "Galilean" satellites in his honor) and Titan show up as disks. Reports come in occasionally that the disk of Rhea or some of the other satellites of Saturn are observed, but these require further verification. The angular diameters can be measured reliably only with disk micrometers, since the satellites have apparent dimensions of only 1" or less.

Surface features are discernible on the largest satellites, although they lie at the limit of resolution for telescopic observations. For example, with a magnification of over 1000 times, the disk of Io measures about 1", and on it dark polar zones and a relatively light, mottled equatorial belt are noticeable. On the disk of Europa, on the other hand, the polar zones are light and the equatorial zone is covered with dark elliptical spots. Ganymede has a similar appearance, with its relatively light polar zones. The surface of Callisto is mostly dark, with isolated lighter spots present against the dark background.

Observations of the apparent migration of spots across the disks of the Galilean satellites show that their rotation periods are equal to the respective periods of revolution about Jupiter, so that all four of these satellites constantly present the same hemisphere to their planet, just as our moon does. On the disk of Titan it is possible to observe limb darkening and some blurred bands, but no conclusions on the rotation of this satellite have been obtained so far.

The motion of a satellite about its primary is associated with various phenomena which are of astrophysical interest to an earthbound observer. The most significant of these phenomena are: occultation of the satellite by the disk of the primary; transit of the satellite across the primary; eclipse of the satellite, that is, its passage through the umbral cone of the primary; and transit of the satellite's shadow across the disk of the planet. These phenomena are regularly predicted and published in astronomical almanacs, and therefore they are systematically observed for the four major Jovian satellites only. In the Saturn system, these events can be observed only during the phase of ring disappearance, and special ephemerides are published on this occasion.

The astrophysical significance of these phenomena is that they make possible an easy comparison of the satellite albedo with the reflectivity of various planetary surface features during the satellite transit. Moreover, observations of the brightness variations when the satellite enters or leaves the shadow of the planet enable us to study the planet's atmosphere by the same methods which are applied to investigate the earth's atmosphere during lunar eclipses. Finally, the time spent by the satellite in the shadow of the planet indicates the satellite diameter.

The photometric studies of satellites are entirely analogous to those carried out for asteroids. They reduce to a determination of the color index and the apparent stellar magnitude m_s by means of stellar photometric

techniques. In addition, by applying formulas (34) and (35) of Chapter XI, we can calculate the stellar magnitude m_0 at mean opposition and the absolute stellar magnitude g. The variations of these quantities with the phase angle indicate the phase coefficient γ, and their periodic fluctuations in time can be regarded as a consequence of the satellite's axial rotation, so that its rotation period can be found.

The geometric and illustrative albedos can be determined for a satellite whose linear radius R is known fairly reliably. The corresponding results are given in Table 3.

TABLE 3

Satellite	δ_1	m_0	g	A_g	A_i	γ	Color	Amplitude of brightness fluctuation	Mass (earth = 1)
Moon	2".40	−12.55	+0.40	0.10	0.15	0.025	G5	—	0.0123
Io	2 .38	+ 5.54	−1.16	0.46	0.69	0.016	G9	0^m.23	0.0121
Europa	2 .08	+ 5.69	−1.01	0.51	0.76	0.019	G5	0 .31	0.0079
Ganymede	3 .62	+ 5.08	−1.62	0.30	0.45	0.023	G3	0 .16	0.0260
Callisto	3 .49	+ 6.28	−0.44	0.11	0.16	0.046	G9	0 .12	0.0160
Titan	2 .9	+ 8.30	−1.26	0.33	0.50	—	—	—	0.0235

δ_1 is the angular semidiameter at a distance of $\Delta = 1$ A.U. The first four columns give Russell's data.

For satellites whose disks cannot be resolved telescopically, the reverse problem can be solved, namely, given a certain geometric albedo, we can find hypothetical values for the radius R. If we solve equation (33) of Chapter XI for R, where A_g is substituted for $\pi\Gamma$, stellar magnitudes are used, and the sun's stellar magnitude is taken as $m_s = -2.5 \lg E_s$, then we obtain

$$\lg R = 0.2m_s - 0.5 \lg A_g - 0.2g.$$

In calculations of R for asteroids and satellites, $A_g = 0.24$ by convention. Hence,

$$\lg R = 3.31 - 0.2g.$$

The hypothetical diameters $2R$ obtained from this formula are listed in the last column of Table 2.

The masses of the largest satellites can be found by analyzing the reciprocal perturbations in the system of the given planet. This requires a large amount of observational material and an elaborate theory describing the satellite motion in the system. Such information is now available for the system of Jupiter and Saturn. Some of the more reliably determined masses are given in Table 3, and we see that Ganymede is the most massive satellite. Among the satellites of the other planets, Triton is found to be quite massive, its mass being close to that of Titan. The masses of the remaining satellites are very small.

Because of their small masses and their correspondingly low escape velocities, most of the satellites should be characterized by a very rapid dissipation of atmospheric gases. Thus, it is to be expected that the

satellites will lack gaseous envelopes. Observations fully corroborate this conclusion. The only exception is Titan, whose spectrum contains a definite methane absorption band near $\lambda = 6190$ Å. The amount of methane is equivalent to a uniform layer 200 m thick. No ammonia has been detected in Titan's spectrum.

REFERENCES FOR CHAPTERS XI AND XII

1. AMBARTSUMYAN, V.A. Teoreticheskaya astrofizika (Theoretical Astrophysics), Chap. VIII. — Gostekhizdat. 1952.
2. BARABASHOV, N.P. Issledovanie fizicheskikh uslovii na Lune i planetakh (Investigation of Physical Conditions on the Moon and the Planets). — Izdatel'stvo Khar'kovskogo universiteta. 1952.
3. BARABASHOV, N.P. Rezul'taty fotometricheskikh issledovanii Luny i planet na astronomicheskoi observatorii Khar'kovskogo universiteta (Results of Lunar and Planetary Photometry at the Khar'kov University Astronomical Observatory). — Khar'kov. 1957.
4. CHANDRASEKHAR, S. Perenos luchistoi energii (Transfer of Radiant Energy). — Izdatel'stvo inostrannoi literatury. 1953 [Russian Translation of "Radiative Transfer," Oxford, Clarendon Press. 1950].
5. KELLOG, W. and K.SAGAN. Atmosfery Marsa, Venery (The Atmospheres of Mars and Venus). — Izdatel'stvo inostrannoi literatury. 1962 [Russian Translation of National Academy of Sciences Edition, Washington, D.C. 1961].
6. KUIPER, G.P., Editor. Atmosfery Zemli i planet (Atmospheres of the Earth and Planets). — Izdatel'stvo inostrannoi literatury. 1951 [Russian Translation from English].
7. LINK, F. Lunnye zatmeniya (Lunar Eclipses). — Izdatel'stvo inostrannoi literatury. 1962 [Russian Translation].
8. MILANKOVIČ, M. Matematicheskaya klimatologiya i astronomicheskaya teoriya klimata (Mathematical Climatology and the Astronomical Theory of Climate). — Gosudarstvennoe ob"edinennoe nauchno-tekhnicheskoe izdatel'stvo. 1939.
9. MOORE, P. Planeta Venera (The Planet Venus). — Izdatel'stvo inostrannoi literatury. 1961 [Russian Translation].
10. Radiolokatsiya Venery (Radar Probing of Venus). — Izdatel'stvo inostrannoi literatury. 1963 [Russian Translation].
11. SHARONOV, V.V. Fotometricheskoe issledovanie prirody planet i sputnikov (Photometric Study of the Nature of Planets and Satellites). — Uspekhi astronomicheskikh nauk, 4:181—249. 1954.
12. SHARONOV, V.V. Priroda planet (The Nature of the Planets). — Fizmatgiz. 1958. *
13. SHIFRIN, K.S. Rasseyanie sveta v mutnoi srede (Scattering of Light in a Turbid Medium). — Gostekhizdat. 1951.
14. SOBOLEV, V.V. Teoriya rasseyaniya sveta v planetnykh atmosferakh (Theory of Light Scattering in Planetary Atmospheres). — Uspekhi astronomicheskikh nauk, 4:250—280. 1954.
15. SUBBOTIN, M.F. Kurs nebesnoi mekhaniki (A Course in Celestial Mechanics), Vol. III. — Gostekhizdat. 1949.
16. SYTINSKAYA, N.N. Absolyutnaya fotometriya protyazhennykh nebesnykh ob"ektov (Absolute Photometry of Extended Celestial Objects). — Izdatel'stvo Leningradskogo Universiteta. 1948.
17. VAUCOULEURS, G. de. Fizika planety Mars (Physics of the Planet Mars). — Izdatel'stvo inostrannoi literatury. 1956 [Russian Translation].

* [English Translation: Israel Program for Scientific Translations, Cat. No. 805. 1965.]

A. V. MARKOV

N 67-15497

Chapter XIII

PHYSICAL PROPERTIES OF THE MINOR PLANETS

§91. Orbits of the minor planets. About 2000 minor planets (also called asteroids) with known orbits have been discovered and recorded so far. The largest of these bodies are never brighter than stars of the 5th or 6th magnitude, while the faintest of the newly discovered asteroids are of the 21st stellar magnitude.

Studies of the orbits of the minor planets have, by international agreement, been allotted to the Institute of Theoretical Astronomy of the USSR Academy of Sciences. These studies have shown that, although the orbits are very diverse, still most of the asteroid trajectories lie in the region between the orbits of Mars and Jupiter. The orbits of all the minor planets are ellipses, with various eccentricities and various inclinations to the plane of the earth's orbit.

The mean distances and eccentricities of some minor planets cause them to go outside of this belt. For example, the orbit of the asteroid Hidalgo (No. 944) has a very high eccentricity and inclination, and its mean distance from the sun is 5.8 astronomical units, that is, greater than the distance of Jupiter. On the other hand, several asteroids have been discovered which pass inside the orbit of Mars. Moreover, there are even a few asteroids, such as Icarus and Adonis, whose perihelia lie inside the orbit of Mercury.

The minor planet Hermes moved in an orbit which, between 25 and 30 October 1937, brought it to within a mere 580,000 km from the earth; thus, this body approached to less than twice the distance of the moon from the earth. It should be noted that the distance of Hermes from the moon was then only 200,000 km.

The total number of minor planets and their relative positions in the solar system are very important for understanding their origin and its relation to the formation of other members of our planetary system. The distribution of the asteroid orbits in space was studied in detail by Hirayama in Japan and by Putilin and Shtaude in the USSR. It was found that sometimes two minor planets move in very close orbits, and also that many orbits almost intersect.

As regards their orbital properties, the minor planets are divided into 5 groups. The first group includes 25 planets which move at an average distance of 3.16 A.U. from the sun. In the second group are 93 planets with an average distance of about 3.05 A.U., in the third group are 15 planets with an average distance of about 2.90 A.U., the fourth group (2.55 A.U.) includes 13 asteroids, and the fifth group (2.25 A.U.) has 57 comparatively close planets in it. Hirayama has suggested that the close similarity of the orbital properties of the planets in each group indicates a common origin for the asteroids in a single group.

The total number of asteroids has been estimated by Putilin as 140,000, counting as an asteroid any body with a diameter of over 1 km. However, the solar system also includes a multitude of even smaller bodies. Numerous meteorites move in interplanetary space, and these sometimes fall onto the earth. The largest meteorites are hundreds of thousands of tons in mass, and their diameters may reach several tens of meters. According to Academician Fesenkov, the meteorites and asteroids all have a common origin.

The asteroids are classified among the minor bodies of the solar system. The other minor bodies are the comets, the meteorites, and the meteoric dust distributed throughout the interplanetary space. Studies of the physical characteristics of the minor planets are of enormous significance in explaining their properties and their origin.

§92. Physical properties of the minor planets. The minor planets, like all the other satellites of the sun, shine by reflected sunlight only. They are so small that with the most powerful telescopes the diameters of just the four largest asteroids could be measured visually with a micrometer or interferometrically (see Table 1). The angular diameters of the planets are given in the table for a distance of 1 A.U. As yet, it is not possible to measure directly the diameters of asteroids fainter than those listed in the table.

TABLE 1

No.	Name	Angular dimension	Diameter, km	Stellar magnitude		Albedo	Color index	Phase coefficient
				absolute	in opposition			
1	Ceres	1".06	767	3.83	7.4	0.10	+0^m.70	0^m.048
2	Pallas	0 .68	489	4.55	8.0	0.13	+0 .36	0 .038
3	Juno	0 .27	193	5.91	8.7	0.22	+0 .52	0 .030
4	Vesta	0 .53	386	3.79	6.5	0.48	—	0 .022

Therefore, to estimate the asteroid size or to study the properties and color of the asteroid surface, we first measure the apparent stellar magnitude; this can be done for all these minor planets, and the measurements are made at various positions relative to the earth and the sun. The ephemerides of the minor planets /2, 3/ published annually by the Institute of Theoretical Astronomy give the following quantities, from which the apparent stellar magnitude m is calculated:

g, the "absolute" stellar magnitude, or the stellar magnitude which the asteroid would have in opposition at a distance of 1 A.U. from the earth and the sun;

m_0, the stellar magnitude of the asteroid at mean opposition, when it is located at a distance of a astronomical units from the sun and $(a-1)$ astronomical units from the earth. Therefore, we may write

$$m = g + 5 \lg r\Delta \quad \text{and} \quad m_0 = g + 5 \lg a(a-1),$$

where r and Δ are the heliocentric and geocentric distances of the asteroids at the given time, expressed in astronomical units.

The data on stellar magnitudes of the minor planets are based on visual photometric measurements and photographic measurements. Reduction of the photographic scale of stellar magnitudes of asteroids to the international system made necessary an average correction of $+0^m.5$ in the range of magnitudes from 7.5 to $9^m.5$ and $1^m.0$ in the range from 9.5 to $13^m.0$.

The observations indicated a distinct dependence of asteroid brightness on the phase angle (that is, on the angle between the directions from the asteroid to the sun and to the earth), the dependence differing for different asteroids (Table 1). The color index C (Table 1) or the difference between the photographic and visual stellar magnitudes of asteroids show a similar dependence.

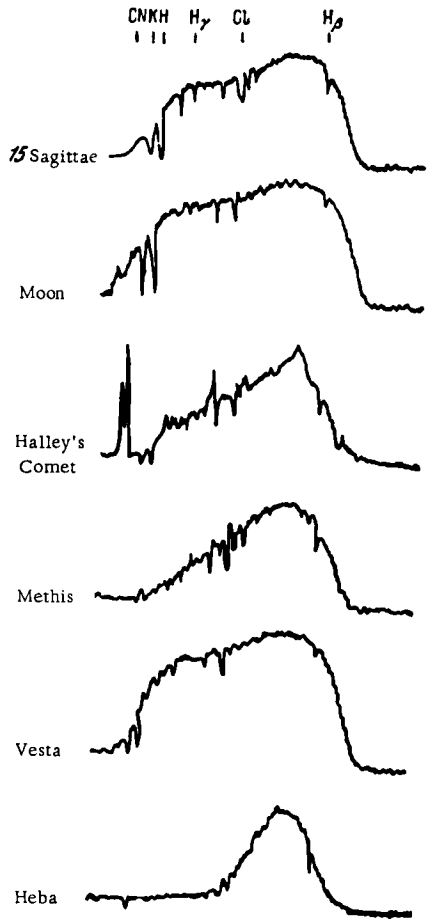


FIGURE 123

The latter quantity makes it possible to estimate to what extent the light reflected from the asteroid is redder or bluer than the illuminating sunlight. The values of C for different asteroids lie in the range from $+0^m.02$ to $1^m.34$.

The spectral reflectivities of some of the minor planets have been studied at Pulkovo and at the Lick Observatory (USA) by comparing the energy distribution in the spectrum of the reflected light with those in the spectra of stars like the sun (Figure 123).

The American results also showed that the asteroid Vesta is similar to the moon in its properties, while the asteroid Methis is similar to the nucleus of Halley's Comet and the asteroid Heba is redder than the moon. In addition, Deich at Pulkovo obtained certain results showing that Vesta is redder than the sun.

Bobrovnikoff /1/ studied the spectra of 12 minor planets brighter than the tenth visual magnitude; he used photographs taken with an 11.4-mm dispersion spectrograph between the H_β and H_ε lines. By comparing these photographs with the spectra of GO stars, he plotted the relative variation of their spectral reflectivity and compared the results with the spectra of the moon and comets. The asteroid spectra were found to be spectra of reflected sunlight, without any additional absorption lines and bands. This indicated that the minor planets have no atmospheres.

However, the surfaces of different asteroids differ in selective-reflection properties. The violet and ultraviolet regions are mostly attenuated, and sometimes they are entirely absent. This refers, in particular, to the spectra of Iris and Heba, while the spectra of Vesta and Pallas (according to Bobrovnikoff) do not differ much from the solar spectrum.

Finally, in cases when the angular diameter of the asteroid and its stellar magnitude are known, we can compute the average reflectivity (or albedo) of its surface (Table 1). We recall that the albedo of a reflecting surface is the ratio of the luminous flux dispersed in all directions to the incident luminous flux. Two mathematical definitions of the albedo are used in connection with the minor planets, namely the Bond (or spherical) albedo /1/ and the geometric albedo. The spherical albedo A_s of a diffusely scattering sphere is the ratio of the luminous flux dispersed by this sphere in all directions to the total luminous flux incident on the sphere.

The geometric albedo is defined as the ratio of the luminous flux dispersed sunward by the planet in opposition (when the phase angle α is zero) to the luminous flux dispersed under the same conditions by a body of the same size, each surface element of which reflects an amount of light equal to that received from the sun in the case of normal incidence. The apparent brightness and the stellar magnitude of a body depend on the phase angle.

The albedos of asteroids have been found to vary widely, from 0.53 to 0.04. In addition, a relation has been observed between the albedo and the rate at which the brightness of the asteroid drops when the body goes away from solar opposition (at which the phase angle is close to 0°) by 1°. The quantity characterizing this rate is the phase coefficient (see Table 1). The higher the albedo of the asteroid, the lower will be its phase coefficient.

The drop in planetary brightness with increasing phase angle depends on the mountain shadows cast on its surface and on the degree of surface roughness. For asteroids, the albedo and surface roughness are related: the higher the albedo, the more pitted is the surface.

Figure 124 shows the phase curve of Ganymede (I) /1/, the theoretical curve for an absolutely mat [orthotropic] sphere (IV) which reflects according to Lambert's law /3/, and the actual phase curves for the moon (II) and

Mercury (III). A comparison of these curves shows that not all the asteroids have surfaces like those of the moon and Mercury, some of them being considerably more pitted than the latter are.

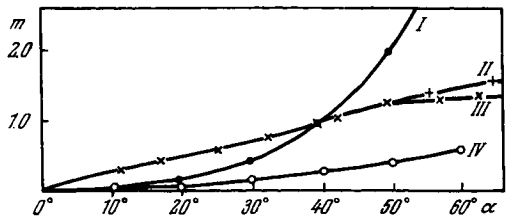


FIGURE 124

Photometric observations of the asteroids have also shown that some have variable brightness, with variation periods of a few hours and amplitudes ranging from a tenth of a stellar magnitude to $1^m.5$ (see Table 2).

TABLE 2

No. of asteroid	Name	Amplitude	Period, hours
5	Astrea	$0^m.20$	3,7
6	Heba	0. 15	3,2
15	Eunomia	0. 4	3,1
39	Leticia	0. 45	2,3
44	Nisa	0. 4	3,2
129	Antigona	0. 4	2,4
340		0. 4	9,2
	Eros	1,5—0,2	5,5

These short-period brightness variations can be attributed either to the fact that the surfaces of the asteroids are unevenly pitted or to the fact that different areas have different albedos. In either case a brightness variation indicates that the asteroid rotates about its axis.

A particular type of brightness variation is observed for the asteroid Eros. The variation period for this body remains fairly constant, but the amplitude may be as much as $1^m.5$ in some epochs and may not exceed $0^m.1$ in other epochs. This phenomenon, which has been proved beyond doubt to exist, can be explained by assuming that this asteroid is double, with two components which revolve rapidly about their common center of gravity. The maximum amplitude of the brightness variation of Eros will thus be observed when the earth lies in the orbital plane of this double system, since at this time one of the components can eclipse the other. Such a conclusion has apparently been confirmed by individual non-Soviet observers, who have reported that through very powerful telescopes they succeeded in resolving the two components of Eros. However, when they tried to calculate the orbital elements of this system, and then the mass and density of its components, they obtained a density of over 45.0 relative to water, which is highly unlikely. Another possible explanation of the variability of Eros

is that its surface is covered by areas with sharply differing albedos. The required brightness curve would then be given by a planetoid whose surface includes three areas with albedos $A_1 = 0.23$ (mica schist), $A_2 = 0.56$ (Etna lava), and $A_3 = 0.08$ (basalt). The variability of the amplitude in this case is explained by the position of the earth relative to the rotation axis of Eros.

Since the sizes and masses cannot be determined directly for all the asteroids, while these data are nevertheless required to establish the relations between the asteroids and other bodies in the solar system and to reach definite conclusions concerning asteroid origins, therefore an indirect method is now used for such calculations. It is assumed that the mean albedo of a given asteroid is equal to the average of the measured albedo values. The mean albedo for the asteroids is $A = 0.24$, and provided the stellar magnitudes of all the asteroids are reduced to a distance of 1A.U., we can calculate their diameters and volumes. (The resulting value of the volume, of course, may be too large or too small by as much as a factor of three.)

The mean density of the asteroids is taken as 3.7 times the density of water. The total volume of all the asteroids taken together is equivalent to a planet 1000 km in diameter, while their total mass is about 1/1000 of the earth's mass. The radii of the smallest asteroids, calculated using these assumptions, are listed in Table 3. For comparison, at the bottom of the table we give the radii of Vesta calculated using its mean and actual albedos. Here, the actual albedo was obtained by measuring the diameter of Vesta with a micrometer.

TABLE 3

Number	Name	Stellar magnitude	Radius, km
1009	Siren	$14^m.4$	2.7
719	Albert	14 .5	2.6
1198	Atlantic	14 .6	2.5
—	Apollo	16 .0	1.0
1121	Amor	17 .0	0.8
	Hermes	14 .4	0.7
	Adonis	18 .5	0.4
4	Vesta	3.47	417.0
		3.79	198.0

§93. The origin of the minor planets. The main reason for studying the minor planets is to establish the part they have played in the cosmogony of the solar system. In order to ascertain this, we must first have some definite knowledge concerning the mass distribution in the asteroid belt and we must also assess the cosmogonic significance of the similarity between asteroids, comets, and meteors.

As mentioned in the last section, the total mass of all the known asteroids is about 1/1000 of the earth's mass (more precisely, from $3 \cdot 10^{-4}$ to $3 \cdot 10^{-3}$ times the earth's mass). It is cosmogonically significant that this figure does not agree with the mass of the asteroid belt calculated from the formulas of celestial mechanics. These formulas take into consideration the perturbations in the motions of nearby planets due to the gravitational attraction of

the asteroid belt. Perturbation theory shows that the total mass concentrated in the asteroid belt should be no less than 1/10 of the earth's mass.

As new asteroids are discovered, the mass total will increase somewhat. However, since it is the smallest asteroids which have not yet been discovered, the total will not rise appreciably, at any rate not enough to agree with the predicted value. Hence, a large part of the matter in the asteroid belt must exist in a finely dispersed form, as meteorites and meteoric dust.

The similarity between the physical properties of meteorites and those of the minor planets has been indicated by the spectroscopic investigations of Krinov, which show that the reflectivities and the color indexes of the meteorites (CI = $+1^m.08$) are close to those of the asteroids. Finally, it is significant, too, that Fesenkov's orbit for the Sikhote-Alin meteorite (which fell on the earth in 1947) is an ellipse whose aphelion lies precisely in the asteroid belt between Mars and Jupiter. Although the mass of this meteorite was of the order of 1000 tons, so that its diameter before entry into the earth's atmosphere could not have been more than 10 m (much less than that of Hermes, whose minimum diameter cannot be less than 500 m), still the orbits of the two bodies were nevertheless close. We may therefore assume that the Sikhote-Alin and Tungusska meteorites, as well as meteor streams, had an origin common to that of the asteroid belt.

Meteorites, and also finer cosmic dust, are located in the asteroid belt and they also travel in orbits which have been distorted into highly eccentric ellipses by the perturbing influence of other objects in the solar system, so that they occasionally hit the earth. These may have been formed by the collision and disintegration of larger asteroids.

The great similarity between the spectra of the nucleus of Halley's Comet and of some asteroids has also led astronomers to suggest that comets and asteroids have a common origin. Some of these scientists maintain that comets are transformed asteroids which, now moving in their elliptical orbits, left the asteroid belt at some time in the past. The appearance of a cometary tail near the sun is generally attributed to the heating of the comet head by the sun.

Now, is it actually possible that the asteroid belt is a product of the disintegration of larger bodies? The following three facts support this hypothesis:

1. A rapid rotation has been observed for some asteroids which have been thoroughly studied by photometric methods and which show short-period variations of brightness.

2. Asteroid orbits can be grouped into families. Some astronomers maintain that each of the orbit families originated due to one of the repeated disintegrations of the fragments of a primordial planet.

3. Meteorites, which may have had the same origin as the asteroids, are (as their structures show) fragments of a previously existing large celestial body (according to Zavaritskii).

Large asteroids such as Ceres, Juno, and Pallas, whose diameters reach hundreds of kilometers, are according to this hypothesis fragments of the primary disintegration which have retained their original size.

4. Finally, the disintegration hypothesis is supported by the fact that some asteroids move in quite similar orbits and that a further fragmentation of the asteroids, leading to the formation of meteorites and meteoric dust, could be the result of collisions between these bodies.

These hypotheses state that the origins of asteroids, meteorites, and comets are related. This theory can be proven only by further studies of these celestial bodies.

REFERENCES FOR CHAPTER XIII

1. PUTILIN, I. I. Malye planety (The Minor Planets). — Gostekhizdat. 1953.
2. SAMOILOVA-YAKHONTOVA, N. S. Malye planety (The Minor Planets). — Uspekhi astronomicheskikh nauk, Vol. 5. 1950.
3. Efemeridy malykh planet (Epheremides of the Minor Planets). — Izdatel'stvo instituta teoreticheskoi astronomiii Akademii nauk SSSR.

O. V. DOBROVOL'SKII

N 67-15498

Chapter XIV

COMETS, METEORS, AND THE ZODIACAL LIGHT

In addition to the major planets and their satellites, the solar system also includes a great number of objects known collectively as "minor bodies." The comets are minor bodies with continually replenished atmospheres. Minor bodies without atmospheres are divided into asteroids, which can be seen through modern telescopes, and meteoric bodies, which are too small to be observed directly through a telescope unless they enter the earth's atmosphere. The light scattered by the accumulation of fine meteoric dust near the ecliptic, and also this accumulation process itself, are called the zodiacal light.

Comets and asteroids are also distinguished by the nature of their orbits: the comets have eccentric orbits, while the orbits of asteroids are close to circular. However, no sharp dividing line can be drawn between comets, meteoric bodies, and asteroids. Asteroids may at times develop atmospheres (Gerta, Elsa, etc.), while comets may for a long time be star-shaped in appearance, without any traces of an atmosphere (Comet Neujmin), or they may even become entirely invisible (Comet de Vico-Swift, with a revolution period of about 6 years, had an atmosphere and was visible only in 1678, 1844, and 1894, that is, on the average once in a century). Finally, some asteroids (Hidalgo, Icarus, Adonis, etc.) have highly eccentric orbits, whereas the orbits of some comets (Schwassmann-Wachmann 1 and Oterma) are almost circular.

§94. General data on comets. A comet usually appears in the form of a faint nebulous spot. Externally a comet is not unlike a globular cluster or nebula, but it can be distinguished from the latter immediately by its rapid proper motion relative to the stars. At the center of the nebulous spot called the head of the comet, a central star-shaped condensation, the photometric nucleus, can often be seen.

A faint telescopic comet may not show a more developed structure. However, when telescopic comets approach the sun, they usually develop tails, which issue from the head or nucleus and extend away from the sun. In addition to the jets which issue from the nucleus and form the tail, the head may have other structural elements, such as halos and envelopes. Halos are formations which are projected onto the celestial sphere in the form of circular arcs or complete circles. The radii of halos increase with time; they are generally concentric with or almost concentric with the photometric nucleus. However, bright comets such as the famous Halley's Comet have several simultaneous halo-formation centers. The envelopes are structural elements in the comet head which are of the same type as the halos. They differ from ordinary halos in their higher brightness and

also in the fact that they invariably have the form of open spheres, which are convex toward the sun and are concentric with or almost concentric with the photometric nucleus.

Cometary tails have been classified by Bredikhin into three types. Type-I tails are formed by jets issuing directly from the photometric nucleus and pointing away from the sun due to solar repulsion. The axis of a type-I tail is approximately parallel to the continuation of the radius vector of the comet. For a faint comet the type-I tail is a narrow, almost rectilinear ray, while for a bright comet the type-I tail is a bundle of rays originating in the nucleus and generally resembling an onion. The individual rays of this bundle may have curves, condensations, waves, or other deviations from the rectilinear form. Figure 125 is a photograph of Comet Morehouse (1908 III), which has a tail of this type.

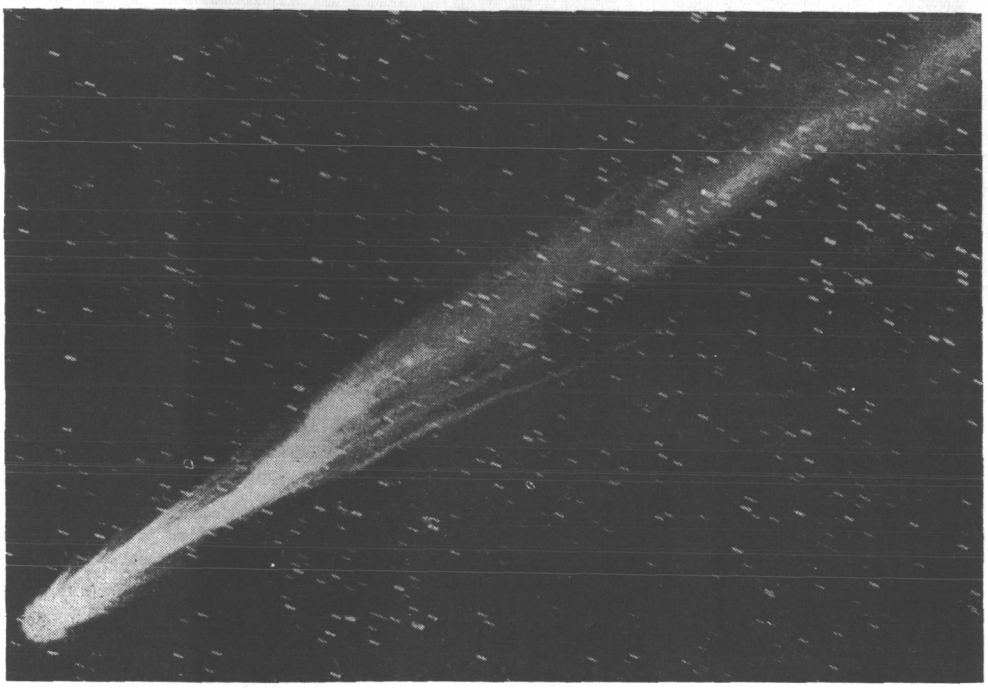


FIGURE 125

A common dynamic feature of all type-I tails is the high acceleration, which may be tens or even hundreds of times greater than the acceleration of the force of gravity toward the sun. This explains why the tail axis is oriented almost parallel to the comet radius vector. Type-I tails also have a common molecular structure: they consist exclusively of ions, mainly Co^+, N_2^+, and Co_2^+.

A type-II tail is an apparent continuation of the comet head. It shows up as a luminous band, which at its origin is as wide as the head and which gradually becomes wider with the distance from the head. The axis of a type-II tail near the comet head is almost rectilinear and points nearly along

the radius vector, like a type-I tail. However, even at a distance of 1° from the head some tail curvature becomes noticeable, the tail deviating toward the direction from which the comet arrived.

Figure 126 is a photograph of Comet Mrkos (1957 V), which has a narrow rectilinear type-I tail and a wide curved type-II tail with numerous synchrones. The accelerations shaping the tail can be determined from its curvature. These were found to be 0.5 to 2.5 of the acceleration of gravity toward the sun.

FIGURE 126

Type-III tails are very similar to type-II tails, but they deviate more from the continuation of the radius vector. Their curvatures correspond to accelerations of 0 to 0.3 times the acceleration of solar attraction.

Type-II and type-III tails apparently are formed mainly from the component matter of the comet head. They consist of dust particles of various sizes; moreover, type-II tails also contain neutral molecules of the head components and products of their decay.

Some comets have anomalous tails which face sunward. According to Bredikhin, these are formed by particularly large particles in which meteor streams originate. The anomalous tails should be distinguished from pseudoanomalous tails, which are ordinary type-III tails whose apparent

sunward directions are due just to the way in which they are projected onto the celestial sphere. Figure 127 is a photograph of Comet Arend-Rowland (1957 II), which has a type-II tail and a pseudoanomalous tail.

Type-II and type-III tails sometimes have structures in the form of rectilinear bands, called synchrones. According to classical concepts, each such band is formed due to the sudden (synchronous) ejection of a cloud of particles of various sizes and thus of various accelerations. We distinguish between complete synchrones, which issue from the nucleus and are characteristic of type-III tails, and terminal synchrones, which originate at some distance from the nucleus and are characteristic of type-II tails (Figure 126).

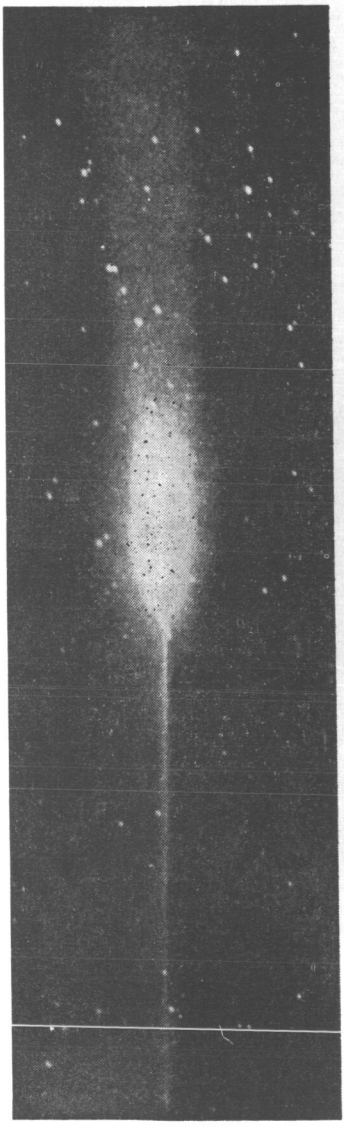


FIGURE 127

The most complete catalog of comets is that of the Royal Astronomical Society (Great Britain), which lists all the comets from 466 B.C. to 1960 whose orbits have been determined. Another well-known catalog is that of Baldet and de Obaldia, which lists comets up to 1952 inclusive (763 apparitions of 540 different comets). For each apparition, the catalog gives the time of perihelion passage T, the revolution period P, the longitude Ω of the ascending node (see Figure 128), the distance ω of the perihelion from the node, the inclination i, the semimajor axis a (for elliptical orbits), the eccentricity, and the perihelion distance q. The comets in this catalog are divided as follows into orbital groups: hyperbolas, 63; $P > 10^3$ years, 353; P between 10^3 and 15 years, 56; $P < 15$ years, 68.

The average annual number of comets observed in the 19th century was 3.1; for the period from 1900 to 1960 it was 5.8; and in the last 30 years an average of 7.1 comets were observed annually. This reflects the continual improvement of astronomical techniques and the organization of systematic observations.

Short-period comets form several "families." The Jupiter family includes 66 comets with $P < 15$ years and aphelion distances close to the aphelion distance of Jupiter (5.5 A.U.); the Saturn family contains 5 comets with P from 13 to 20 years, all having aphelion distances close to that of Saturn (10 A.U.); the Uranus family has 6 comets with P from 27 to 50 years and aphelia close to that of Uranus (20 A.U.); and the Neptune family includes 10 comets with P from 50 to 100 years, all having Neptune's

TABLE 1

Comet	P	ω	Ω	i	a	q	N	T	H_{10}
1960 i Encke	$3^a.30$	185°.2	334°.7	12°.4	2,22	0.338	46	1961 Feb 12,8	$12^m.9$
1961 g Grigg-Skjellerup	4 .90	356 .3	215 .4	17 .6	2,89	0.855	10	Dec 31.4	13. .4
1954 III Honda-Mrkos-Pajdušáková	5 .21	184 .1	233 .1	13 .2	3,01	0,556	2	Feb 5,1	12. .7
1957 II Tempel 2	5 .27	191 .0	119 .3	12 .5	3,03	1,369	12	Feb 4.9	13 .0
1927 I Neujmin 2	5 .43	194 .7	328 .0	10 .6	3,09	1,338	2	Jan 16.2	10 .7—11.9
1879 I Brorsen [1]	5 .46	14 .9	101 .3	29.4	3.10	0,590	5	March 31.0	9 .3
1962 b Tuttle-Giacobini-Kresák	5 .49	38 .0	165 .6	13 .8	3,11	1,122	4	Apr 23.9	12 .0
1908 II Tempel 3-Swift	5 .68	113 .7	290 .3	5 .4	3,18	1,153	4	Oct 1.4	12 .8
1894 IV de Vico-Swift	5 .86	296 .6	48 .8	3 .0	3,25	1,392	3	Oct 12.7	10 .2
1879 III Tempel 1	5 .98	159 .5	78 .8	9 .8	3,30	1,771	3	May 7.6	10 .4
1951 VI Pons-Winnecke	6 .12	170 .2	94 .4	21 .7	3,35	1,161	15	Sept 8.6	14 .1
1958 I Kopff	6 .32	161 .7	121 .0	4 .7	3,37	1,516	8	Jan 20,1	11 .2
1961 a Forbes	6 .42	259 .7	25 .4	4 .6	3,46	1,545	4	July 24.8	13 .0
1959 VII Giacobini-Zinner	6 .42	172 .8	196 .0	30 .9	3,46	0,936	7	Oct 26.9	11 .8
1958 V Wolf-Harrington	6 .52	187 .0	254 .3	18 .5	3,48	1,605	2	Aug 11.7	
1960 j Schwassmann-Wachmann 2	6 .53	357 .9	126 .0	3 .7	3,49	2,150	6	Sept 5	10 .3
1852 III Biela	6 .62	223 .3	245 .9	12 .6	3,53	0,861	6	Sept 24,2	8 .1
1950 V Daniel	6 .66	7 .2	69 .7	19 .7	4,54	1,465	4	Aug 24,3	12 .8
1954 XI Wirtanen	6 .68	343 .5	86 .5	13 .4	3,60	1,626	2	Aug 13,5	14 .7
1950 II D'Arrest	6 .70	174 .4	143 .6	18 .1	3,55	1,378	10	June 6.6	11 .7
1957 VII Arend-Rigaux	6 .71	326 .4	124 .6	[redacted] .2	3,55	1,385	2	Sept 8.7	
1961 h Perrine-Mrkos	6 .71	165 .9	240 .3	17 .7	3,55	1,150	4	1962 Feb 11	9 .1
1960 IX Reinmuth 2	6 .71	45 .4	296 .2	7 .0	3,55	1,933	3	Nov 24.9	12 .2
1960 VI Brooks 2	6 .72	197 .1	176 .5	5 .5	3,66	1,763	10	June 17,3	13 .7
1960 VII Harrington	6 .80	232 .8	119 .2	8 .7	3,48	1,582	3	June 29.0	
1906 III Holmes	6 .86	14 .3	331 .7	20 .8	3,61	2,122	3	March 14,6	9 .8
1956 V Johnson	6 .87	205 .9	118 .2	13 .9	3,61	2,259	2	July 24,2	10 .0
1960 VIII Finlay	6 .90	321 .6	42 .1	3 .6	3,62	1.080	7	Sept 1,1	11 .6
1960 V Borrelly	7 .02	350 .9	76 .2	31 .1	3,66	1,449	7	June 12.6	12 .5
1962 a Harrington-Ebell	7 .24	338 .3	145 .9	16 .8		1,785	2		
1961 c Faye	7 .38	199 .1	203 .6	9 .1	3,79	1,610	15	May 16,0	11 .1
1962 f Whipple	7 .46	190 .0	188 .4	10 .2	3,81	2,471	5	1963 Apr 29,6	11
1956 II Ashbrook-Jackson	7 .50	349 .1	2 .3	12 .5	3,83	2,324	2	Apr 5,67	0 - 10.0
1958 II Reinmuth	7 .65	12 .9	123 .6	8 .4	3,83	2,026	4	March 25,0	12 .4
1959 V Arend	7 .80	44 .6	357 .6	21 .7	3,93	1,83	2	Sept 1,2	
1958 IV Oterma	7 .92	354 .6	155 .1	4 .0	3,97	3,406	—	June 10,5	7 .2
1960 III Schaumasse	8 .17	51 .8	86 .4	12 .0	4,06	1,194	5	Apr 18,0	9—10,3
1959 II Wolf 1	8 .43	161 .1	203 .9	27 .3	4,12	2,507	10	March 21,8	13 .2
1960 f Comas Solá	8 .59	40 .0	62 .8	13 .4	4,19	1,777	5	1961 Apr 4,5	9 .7
1960 IV Väisälä 1	10 .46	44 .4	135 .4	11 .3	4,79	1,741	3	May 10,8	13 .4
1951 V Neujmin 3	10 .95	144 .8	156 .2	3 .8	4,93	2,032	2	May 28,4	13 .7
1933 I Gale*	10 .99	209 .1	67 .3	11 .7	4,94	1,183	2	June 18,5	9 .2
1939 X Tuttle 1	13 .61	207 .0	269 .8	54 .7	5,70	1,022	8	Nov 10,1	11 .4
1925 II Schwassmann-Wachmann 1	16 .10	355 .8	321 .6	9 .5	6,39	5,538		1957 May 12,9	2 .5
1948 XIII Neujmin 1	17 .93	346 .7	347 .1	15 .0	6,85	1,545	3	Dec 15,9	11 .9
1956 VI Crommelin	27 .87	196 .0	250 .4	28 .9	9,19	0,743	5	Oct 19,4	10 .7

* [This should be 1938 I Gale (Translator)]

TABLE 1 (continued)

Comet	P	ω	Ω	i	a	q	N	T	H_{10}
1942 IX Stephan-Oterma	$38^{a}.96$	358° .3	78° .5	17° .9	11.41	1.596	2	Dec 19.2	$9^{m}.5$
1913 VI Westphal	61 .73	57 .1	346 .8	40 .9	15.62	1.254	2	Nov 26.8	8 .8
1919 III Brorsen-Metcalf	69 .06	129 .5	310 .8	19 .2	16.83	0.485	2	Oct 17.4	9 .6
1956 IV Olbers	69 .57	64 .6	85 .4	44 .6	16.91	1.179	3	June 15.9	5 .5
1954 VII Pons-Brooks	70 .88	199 .0	255 .2	74 .2	17.13	0.774	3	May 23.0	6 .1
1910 II Halley	76 .03	111 .7	57 .3	162.2	18.25	0.587	29	Apr 20.2	4 .6
1939 VI Herschel-Rigollet	156.0	29 .3	355 .1	64 .2	28.98	0.748	2	Aug 9.5	8 .5
1907 II Grigg-Mellish	164.3	328 .4	189 .2	109.8	30.00	0.923	2	March 27.7	10 .4

aphelion (30 A. U.) (numerical data according to Vsekhsvyatskii). All the comets of the Jupiter family possess direct motions and have very small inclinations. The other families include both direct and retrograde comets. Long period and aperiodic comets have very different values of i.

All the periodic comets which were observed at least twice prior to 1 January 1963 are listed in Table 1.

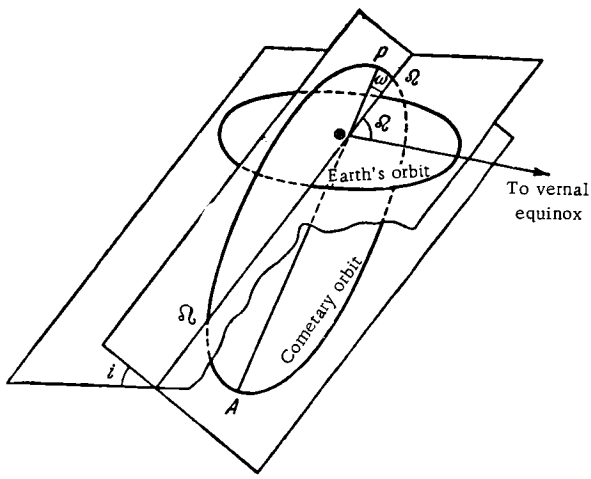


FIGURE 128

Many empirical formulas have been proposed to describe the systematic variation of cometary brightness with the heliocentric distance r. The generally accepted formula is that of S. V. Orlov:

$$H = m - 5 \lg \varrho - 2.5n \lg r, \qquad (1)$$

where ϱ is the heliocentric distance of the comet, m is its apparent stellar magnitude, and H and n are photometric parameters which remain constant for a given series of observations. H is usually called the "absolute magnitude" of the comet, and n characterizes the rate of brightening as the comet approaches the sun.

Vsekhsvyatskii has compiled a catalog of the photometric parameters of comets, which includes all the comets in the catalog of Baldet and de Obaldia and also the apparitions of 1953—1955; in addition, regular supplements to this catalog are published. The mean value of n according to Vsekhsvyatskii is about 4. Correspondingly, he has proposed an alternative definition of the absolute magnitude; according to this definition, the absolute magnitude H_{10} of a comet is the parameter H determined for $y = 2.5n = 10$ (see formula (1)). In other words, Vsekhsvyatskii's absolute magnitude of a comet is the apparent magnitude which the comet would have at unit distances from the earth and the sun if the rate of brightening of the comet were exactly 10.

Table 2 gives the distribution of comets in terms of Vsekhsvyatskii's absolute magnitude H_{10}. The table gives values for the direct observations of comets of 5th to 7th magnitudes which are far too low, due to the low probability of detecting faint comets, the actual number of which is much greater.

TABLE 2

Range of absolute magnitudes	Number of comets	Number of comets, corrected for visibility conditions
-3^m to -2^m	1	
−2 to −1	1	
−1 to 0	1	1
0 to 1	4	4
1 to 2	5	5
2 to 3	12	16
3 to 4	23	42
4 to 5	52	130
5 to 6	70	236
6 to 7	77	349
7 to 8	58	430
8 to 9	41	465
9 to 10	35	521
10 to 11	20	760
11 to 12	7	1240
12 to 13	3	
13 to 14	1	

It should be noted that the two distribution functions of Table 2 are only tentative. A single comet usually has different H and n during different apparitions and even in different parts of its trajectory (for example, before and after perihelion). This may indicate a systematic dependence of brightness estimates for comets on the visibility conditions. Comets also undergo a secular loss of brightness, which varies from a few hundredths to one or one and a half stellar magnitudes per revolution.

§95. Cometary nuclei. The masses M_α of cometary nuclei are determined the most directly from their gravitational effects. However, since these effects are weak, only the order of magnitude of M_α can be found. When Comet Lexell approached the earth, its period became 2.5 days shorter, whereas the length of the year did not change at all. If the year had changed by 1 sec, then M_α would have been $1.3 \cdot 10^{-4} M_\delta$. The period of Comet Brooks changed from 29 years to 7 years due to its approach to Jupiter in 1886, while the period of Jupiter did not change noticeably. Thus we know that $M_\alpha \ll 10^{-4} M_\delta$. Comet Biela divided into two parts, and then the mutual

perturbations of the parts indicated that $M_{\text{ядра}} \approx 4.2 \cdot 10^{-7} M_{\odot}$. Finally, Comet Wirtanen (1957 VI) ejected a secondary nucleus which had a velocity of ~ 1.6 m/sec. Assuming that this value represents the parabolic velocity of the nucleus and that the point of separation was 10 km away from the center of the nucleus, we obtain $M_{\text{ядра}} \approx 10^{17}$ grams.

A partial solution of the three-body problem gives the following condition for stability of a cometary nucleus relative to the tidal forces of the sun: $R/q < \frac{1}{3}(M_{\text{ядра}}/M_{\odot})^{1/3}$. If we assume for the radius of the cometary nucleus the high value of R = 25 km, then we obtain $M_{\text{ядра}} > 0.5 \cdot 10^{-14} M_{\odot}$. These estimates are consistent with the upper limits for the masses obtained from direct observations. Cometary nuclei are too small to be resolved with a telescope. Therefore, by taking 0".3 as the diameter of the nucleus and by taking a density of 6 g/cm^3, Richter obtained upper limits of $M_{\text{ядра}}$ ranging from $1.2 \cdot 10^{-5} M_{\odot}$, or $7.3 \cdot 10^{22}$ g (comet 1882 I), to $5.5 \cdot 10^{-11} M_{\odot}$, or $3.3 \cdot 10^{17}$ g (Comet Neujmin 1).

According to classical concepts dating back to Laplace and Bessel, the nucleus consists of low-melting substances in the form of ice or snow. Later, however, the discovery that comets and meteor streams are related gave rise to the stony-nucleus model, which has been considered in detail by Levin. According to Levin, volatile substances are present in the stony nucleus in the form of sorbed gases, and the development of the atmosphere is interpreted as a desorption of gases from the surface. The comet brightness is $i \sim \exp\left(-\frac{E}{RT_0}\right)$, where E is the heat of desorption, T_0 is the temperature of the nucleus surface, and R is the universal gas constant. Rijves improved this model by taking into consideration the diffusion of sorbed gases from the interior of the nucleus to the surface; he interpreted E as the activation energy of diffusion. Then, $T_0 \sim r^{-\frac{1}{2}}$ (see below), and

$$i \sim \exp\left(-\frac{\text{const}}{\sqrt{r}}\right) \tag{2}$$

or, in stellar magnitudes,

$$m = A + B\sqrt{r}, \tag{3}$$

where A and B are constants. This is Levin's formula, which in practice gives the same variation of m with r as (1) does; it was checked and found valid by Levin, by Oort and Schmidt, by Vorontsov-Vel'yaminov, and by many others.

However, although the "stony" model gives an accurate dependence of i on r, still it does not explain by what mechanism the dust component appears in cometary atmospheres. This difficulty, on the other hand, is overcome by combining the "ice" model and the "stony" model.

For example, Whipple's synthesis of the two models regards the nucleus as a conglomerate of stony fragments and ices with various chemical compositions (a mixture of "contaminated" ices). It is assumed here that the surface layer contains a high proportion of the high-melting-point stony component, which forms a protective layer through which the volatile substances diffuse. Vsekhsvyatskii's model assumes a stony nucleus partly covered with ice.

Whipple's model has been further developed by Swings and Delsemme, who noticed that the partial saturated-vapor pressure p is very different for different "parent" ices, so that the rates of vacuum sublimation should also differ greatly, a situation which is actually not observed. The "parent" molecules therefore must exist not in the form of pure ices, but rather as frozen hydrates of the compound $M \cdot nH_2O$, where M is the corresponding parent molecule and nH_2O is a certain number of water molecules which depends on the chemical properties of the molecule M. Such compounds may form by the accretion of molecules in a vacuum. These hydrates all have similar physical properties; in particular, they all have almost the same heat of vaporization, a value close to that for pure water, so that these frozen compounds sublimate at the same rate as ordinary H_2O ice.

The above inconsistency, however, could also be explained within the framework of the "ice-mixture" model, provided a sufficiently fine dispersion of the mixture is assumed. Then, the amount of chemically different gases liberated will be proportional to the vaporized amount of the highest-melting-point component; this highest-melting component will be H_2O ice, as in the Swings-Delsemme model.

The "ice" model of Donn and Urey assumes that the free radicals observed in cometary atmospheres are present in a free state in the solid nucleus also, where they could appear due to freezing of interplanetary gases consisting of these radicals. If their concentrations are not too high, the radicals may exist in stable form at low temperatures.

Actual cometary nuclei probably include to some degree all the different structures mentioned above: mechanical mixtures of different dispersities, hydrates, and chemically-active radicals. Also, the surface layer may be partly or completely covered by a high-melting-point crust through which the volatile components diffuse.

Possible mechanisms by which matter can leak out of these conglomerate nuclei have been considered by various authors, primarily by Whipple, Weigert, and others. If the surface is abundantly covered with free ices and the protective crust has not yet formed (a "young" comet which has completed only a small number of perihelion passages), then the sublimation is very vigorous. The gases here entrain dust particles from the surface of the nucleus. The spectrum of the comet is therefore predominantly one of reflected sunlight (as in 1957 II Arend-Rowland and 1957 V Mrkos), and this serves as a distinguishing characteristic between "young" and "old" comets (another distinguishing feature is the nature of the orbit).

In "old" comets the gases escape through a solid crust which covers the nucleus; the kinetic energy of the gas is here not high enough to overcome the cohesive forces between the dust particles of the crust. Thus, this dust apparently is found in the cometary atmosphere only if the crust is damaged somehow. The spectra of comets are mainly gas spectra, in which the solar continuum shows up weakly or is missing altogether (a typical example is Comet Encke).

The thermal regime of the surface layer of the nucleus, which determines the rate of gas liberation (and therefore the brightness of the comet as well), is in general found by solving the heat-conduction equation:

$$\delta c \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(K \frac{\partial T}{\partial x}\right) \qquad (4)$$

with the boundary condition

$$(1-a)\frac{q}{r^2} = -K(\text{grad}\,T)_0 + \sigma T^4 + b, \tag{5}$$

where t is time, x is the normal to the surface of the nucleus, K is the coefficient of heat conduction, δ is the density and c is the specific heat of the surface layer of the nucleus, q/r^2 is the thermal-energy flux incident from the sun per cm^2 of nucleus surface, a is the albedo, $-K(\text{grad}\,T)_0$ is the heat influx to the interior of the nucleus, σT^4 is the radiation loss, and b is the thermal energy consumed during evaporation per unit surface per unit time. The last quantity is related to the temperature T_0 of the vaporizing surface and to the saturated-vapor density p at this temperature by the expressions

$$\left.\begin{aligned} Q &= \frac{b}{L} = \frac{\alpha p}{\sqrt{2\pi m_m k T_0}}, \\ p &= A T_0^{(c_p - c)/k} e^{-\frac{L}{KT_0}}. \end{aligned}\right\} \tag{6}$$

Here $L = (c_p - c)\,T_0 + L_0$ is the latent heat of vaporization of a single particle at the temperature T_0, L_0 is the heat of vaporization of a single particle at 0°K, c_p is the specific heat of vapor at constant pressure, m_m is the mass of the particle (molecular mass), A is a constant, Q is the number of particles vaporizing per unit area per unit time, and α is the vaporization coefficient.

A numerical integration of equation (4), taking into account (5) and (6) as well, was carried out by Minnaert and Markowitz for various values of the thermal constants. In particular, Minnaert found that a cometary nucleus with a diameter of 1 km moving in the orbit of Halley's Comet will have a surface temperature of about 50°K at aphelion and about 33°K at perihelion; the variation of the temperature is asymmetric about the perihelion (Figure 129). Markowitz found that the results of the numerical integration could always be approximated by the expression $T_0 = T_{00} r^{-\alpha_0}$, where $\alpha_0 \leqslant 0.5$. By substituting this expression for T_0 into (6) and neglecting the variation of $T^{(c_p - c)/k - 1/2}$ relative to the exponential function, we find for the rate of gas separation from the nucleus

$$\lg Q = C_1 + C_2 r^{\alpha_0}.$$

However, the brightness of the comet is mainly determined by the quantity Q, since the other two factors, namely the lifetimes of the luminous molecules of the head until dissociation ($\sim 1/r^2$) and the frequency of excitations by solar photons ($\sim r^2$), mutually cancel one another. Thus, we obtain Levin's generalized formula

$$m = A + Br^{\alpha_0} \quad \left(\alpha_0 \leqslant \frac{1}{2}\right), \tag{3'}$$

first derived empirically by Vsekhsvyatskii.

The formation of a dust atmosphere around conglomerate nuclei has been explained as follows by Weigert. First, we write the equation of motion for a meteoric particle of radius h, density δ, and mass $m = \frac{4}{3}\pi\delta h^3$ in the form

$$m\ddot{x} = \dot{P}(x) - K(x),$$

where $\dot{P}(x) = 2m_m(v - x)\,n(x)$ is the momentum increase due to collisions with vaporizing gas molecules, $n(x) = \pi h^2 Q \frac{R_0^2}{x}\frac{v-x}{v}$, $K(x) = \frac{k^2 M_\odot m}{x^2}$ is the force of gravitational attraction of the nucleus, v is the velocity of the gas molecules, R_0 is the radius of the nucleus, and x is the distance from the center of the nucleus. Then, for the mean density ϱ of the nucleus, we may write

$$x^2\ddot{x} + \alpha\dot{x} + \beta\dot{x}^2 + \gamma = 0,$$

where $\alpha = \frac{3m_m R_0^2}{\delta}\frac{Q}{h}$, $\beta = \frac{v}{2a}$, and $\gamma = \frac{4}{3}\pi k^2 \varrho R_0^3 - \frac{1}{2}\alpha v = \gamma_0 - \frac{1}{2}\alpha v$. When the velocity is maximum, $\ddot{x} = 0$ and we have

$$v_{max} = v - \sqrt{\frac{\gamma_0}{\beta}}. \tag{7}$$

At the beginning of motion, $\ddot{x} \gg 0$, that is, $\gamma \leqslant 0$, so that the radii of the largest particles entrained by the gas stream are

$$h_m = \frac{9}{\sqrt{8\pi^3}}\frac{Q\sqrt{m_m k T_0}}{k^2 \varrho \delta R_0}. \tag{8}$$

For example, in a nucleus with a radius of 5 km where NH_3 is vaporized (surface temperature, T_0 = 120°), the gas will entrain all particles with radii $\leqslant 10^{-2}$ cm.

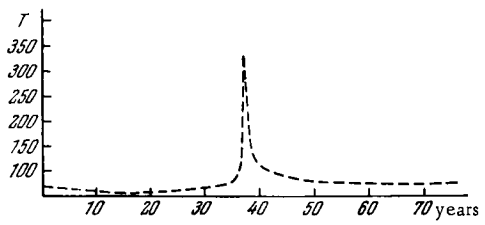


FIGURE 129

In "old" comets with large r, when b and σT^4 are small, the conditions are those of the classical problem of heat-conduction theory. For periodic heating of the surface of a rotating nucleus, when a sufficiently long time has elapsed from the moment when heating began, the solution is independent of the initial conditions. If the surface temperature is given by the series

$$T(0, t) = \sum_0^\infty T_n \cos(n\omega t + \varepsilon_n) \tag{9}$$

with known coefficients T_n and ε_n and a cyclic rotation frequency of the nucleus $\omega = \frac{2\pi}{P}$, where the thermal diffusivity $k = \frac{K}{\delta c}$ is constant, then the solution has the form

$$T(x, t) = \sum_0^\infty T_n e^{-x\sqrt{\frac{\pi n}{Pk}}} \cos\left(\frac{2\pi n}{P}t - x\sqrt{\frac{\pi n}{Pk}} + \varepsilon_n\right). \tag{10}$$

From this solution we can draw some general conclusions which are independent of the numerical values of T_n and ε_n. For instance, the quantity $\frac{1}{2\pi}\sqrt{\frac{Pk}{\pi n}}$ is the wavelength of the nth thermal wave, which has a period $\frac{P}{n}$ The phase velocity of the thermal wave which determines the propagation velocity of some particular phase, such as the temperature maximum, is

$$v_{ph}=2\sqrt{\frac{\pi kn}{P}}. \tag{11}$$

The amplitude of temperature variation at a depth x is $1/e^{x}\sqrt{\frac{\pi n}{Pk}}$ times the amplitude at the surface; in particular, at a depth

$$x_e=\sqrt{\frac{Pk}{\pi n}} \tag{12}$$

the amplitude is $1/e$ times the surface amplitude. For example, with a period of 24 hours ($\sim 10^5$ sec) and a thermal diffusivity $k = 3\cdot 10^{-5}$ cm^2/sec, the nucleus can be regarded as quite isothermal even at a depth of 1 cm.

The heat flux into the interior of the nucleus, defined as

$$F=-K\left(\frac{\partial T}{\partial x}\right)_0, \tag{13}$$

has a phase shift of $\frac{P}{8}$ relative to $T(0, t)$. The temperature $T(0, t)$ may possess phase shifts from 0 to $\frac{P}{8}$ relative to q. The lower value corresponds to high radiation losses, while the upper value corresponds to low radiation and sublimation losses and to $F = q$.

For "old" comets with small r the radiation losses

$$T_0=\left(\frac{q_{eff}}{\sigma}\right)^{1/4}\frac{1}{\sqrt{r}} \tag{14}$$

prevail. The numerical value of the constant $\left(\frac{q_{eff}}{\sigma}\right)^{1/4}$ depends on the size of of the body, the albedo of the surface, and the angle of incidence of solar radiation. For normal incidence at the perfectly black surface of a massive nucleus, $q_{eff} = q_0$, and the constant is equal to 330°K. Perfectly black meteoric dust particles (radius < 1 cm) which are heated throughout their volumes will, according to Zanstra, have $T_0=\frac{289\,^\circ\mathrm{K}}{\sqrt{r}}$.

§96. Cometary spectra. In the 19th century, observers (Donati, Schiaparelli, Bredikhin, and others) established that cometary spectra are made up of a continuous background with emission bands (mainly molecular) superimposed onto it.

Continuous spectra are characteristic of the central part of the comet head and of type-II and type-III tails. According to Shul'man and others, the energy distribution in the continuum is like that in the solar spectrum, with a maximum near 4700 Å; the usual Fraunhofer lines of the solar spectrum are also visible in the brightest parts of the comet-head continuum. Therefore, the continuous spectrum is due to the scattering of sunlight by dust particles in the comet. This conclusion is doubtless true for the spectrum of the comet head and is also quite probable for the spectra of type-II and type-III tails. The polarizations and colors of tails noted by various observers also indicate this. The relative significance of the

various forms of scattering in comets was studied by Poloskov and, in great detail, by Oiseau and Batteau, who showed that Mie-type scattering at small dust particles with diameters of 0.2 to 0.5μ is the most probable.

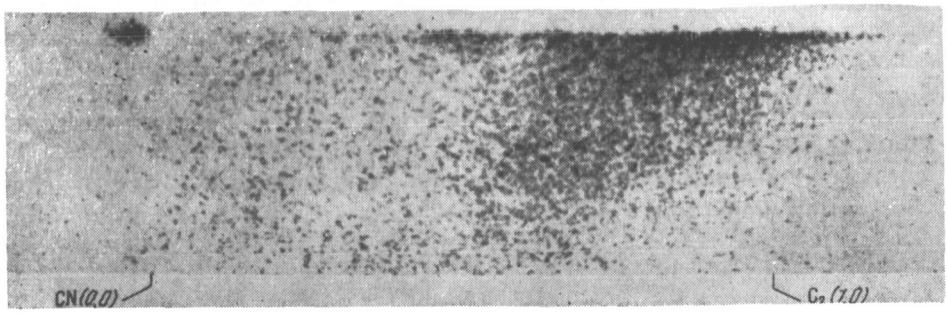


FIGURE 130a

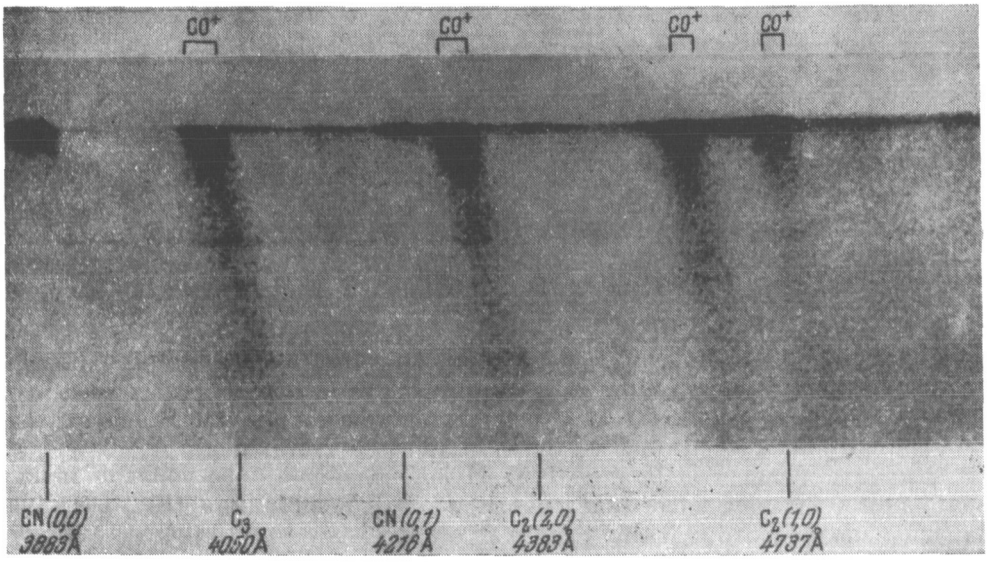


FIGURE 130b

Around the turn of the century, some of the molecular emissions were identified as C_2, CH, and CN bands for comet heads, and as N_2^+ and CO^+ bands for type-I tails. Figure 130 shows two cometary spectra photographed with a prism objective: Figure 130 is the spectrum of Comet 1910 I, which had a type-II tail (the continuous spectrum is predominant); Figure 130b is the spectrum of Comet 1911 V, which had a type-I tail (CO^+ emission predominant). Figure 131 shows slit spectrograms of the head of Comet Encke (1947 XI) (here, a and b were photographed with a glass prism, and c with a quartz prism). Subsequently, C_3, NH, OH, NH_2, and OH^+ were identified in the head, CO_2^+ in the tail, and CH^+ in both the head and the tail. A list of the principal emission bands for the head is given in Table 3, and for the tail in Table 4. The atomic lines observed were those of Na, O, Mg, Fe, and Ni. The last three elements were observed only at very small

distances from the sun, of the order of 10^{-2} A.U. (and only the Fe multiplets were established with any certainty). Na appears when $r \leqslant 0.7$ A.U., and it shows up as the familiar $3^2P - 3^2S$ doublet near λ 5890 — 5896 Å; this is the reason for the characteristic yellow coloring of the heads and tails of bright comets. Emissions of atomic oxygen have been found only recently, and they show up as weak forbidden lines in the green and red regions. For the Comet Arend-Rowland (1957 II) a transition was observed between the sublevels of the lambda-doublet of the level $J'' = 1.5$ of the state $^2\Pi_{3/2}$ in the radio range (frequency, 600 Mc).

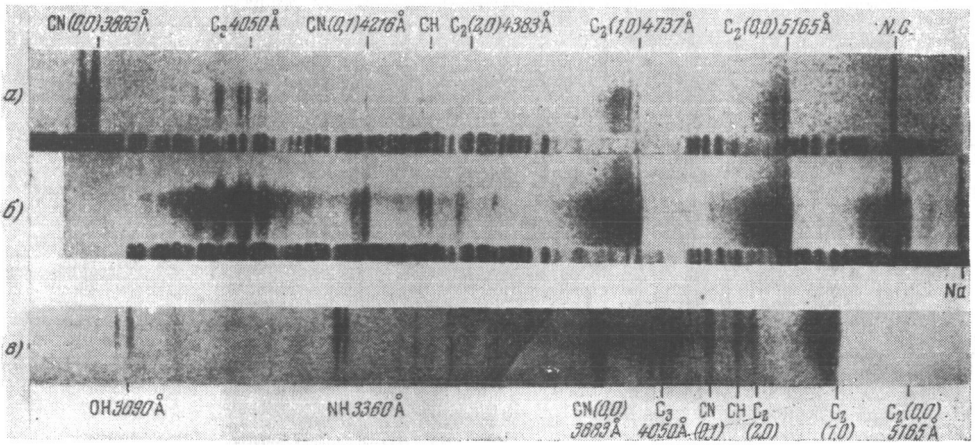


FIGURE 131

Spectra of diatomic molecules are thus of very great significance in cometary spectroscopy. Consequently, let us first consider the spectroscopic properties of these molecules. A diatomic molecule always has an axis of symmetry, namely the straight line joining the two nuclei. The electron state is therefore characterized primarily by the projection L_z of the total orbital momentum L of the electrons onto this axis. In accordance with the general quantization rules for mechanical momenta and their projections, we have

$$L_z = \Lambda \frac{h}{2\pi}, \quad |L| = \sqrt{L(L+1)}\,\frac{h}{2\pi}, \tag{15}$$

where the quantum number Λ may take the values $0, \pm 1, \pm 2, \pm 3, \ldots$ The corresponding electron states are designated by the Greek letters, $\Sigma, \Pi, \Delta, \Phi, \ldots$ No states higher than Δ are observed in cometary spectra. The quantum number S of the total spin of the electrons determines the multiplicity of the state, that is, the number of component levels into which the given state breaks up in an external [magnetic] field. The multiplicity is, just as for atoms, equal to $2S + 1$, and it is designated by a superscript to the left of the term symbol (for example, the state with $\Lambda = S = 0$ is designated as $^1\Sigma$).

A more detailed description of the electron state can be achieved by introducing the vectors N (the angular momentum of the nuclei about the center of gravity) and J (the total angular momentum). The vector J is

TABLE 3

Molecule	System (electron transition)	Vibrational transition	λλ in Å	λ_{max}	Remarks
C_2	$A^2\Pi_g - X^2\Pi_u$ (Swan system)	$\Delta v = +2$	4364.3 — 4380.9	4380	System determines visual diameter of head
		$\Delta v = +1$	4669.5 — 4734.9	4713	
		$\Delta v = 0$	5097.3 — 5164.9	5164	Indicates a high rotational and vibrational temperature, similar to that observed under laboratory conditions (~ 3000°K), which corresponds to a homonuclear molecule.
		$\Delta v = -1$	5466.2 — 5634.1	5634	
		$\Delta v = -2$	5959.6 — 6186.5	6186	
CN	$B^2\Sigma^+ - X^2\Sigma^+$ (violet system)	1 — 0	3572.2 — 3584.3	3584	System determines photographic diameter of head. Very pronounced characteristics of rotational structure, caused by presence of Fraunhofer lines in solar exciting radiation; low rotational temperature, corresponding to heteronuclear molecule.
		0 — 0	3851.0 — 3883.0	3882	
		0 — 0	4193.0 — 4214.7	4214	
C_3			3950.2 — 4137.8	4051	Exists in carbon stars of Class N and in planets.
OH	$A^2\Sigma^+ - X^2\Pi_i$	0 — 0	3078.7 — 3107	3090	Anomalous intensity distribution, due to Fraunhofer lines of solar exciting radiation; low rotational temperatures (heteronuclear molecule).
		1 — 1	3135.0 — 3159	3140	
NH	$A^3\Pi_i - X^3\Sigma^-$	0 — 0	3350.8 — 3372.0	3358	Low rotational temperatures
CH*	$B^2\Sigma^- - X^2\Pi$	0 — 0	3886.6 — 3921.5	3890	Q and P branches
	$A^2\Delta - X^2\Pi$	0 — 0	4281 — 4348	4313	Q, R and P branches
CN	$A^2\Pi - X^2\Sigma^+$ (red system)	2 — 0	7858 — 7957	7906	
		3 — 1	8046 — 8164	8106	
NH_2	α band of ammonia			5384	Observed near nucleus. Extends over entire visible spectrum; in several cases superimposed onto emission of other molecules, for example, C_2 at $\lambda = 5166$ Å.
				5428	
				5706	
				5978	
				6299	

different, depending on which two of the vectors L, S, and N have the strongest coupling. Four coupling modes are distinguished (a, b, c and d according to Gund), plus some intermediate cases; cases a and b are the most frequent. In case a, the vectors L and S are first added, giving a total axial projection with a quantum number $\Omega = L + S$. This projection* is then added to the vector N, which is perpendicular to it, giving J. The axial projection of J has a quantum number J, which may take the values

* Vectors L and S precess rapidly about the axis; therefore, in the following we will consider only the projection onto the axis which remains constant in time.

Ω, $\Omega+1$, . . . The numerical value of Ω is indicated in this case by a subscript to the right of the term symbol. In case *b*, S is not coupled with the internuclear axis, and Ω has no physical meaning. Here Λ and N have a resultant K (the total momentum without spin). The axial component of K is determined by the quantum number $K = \Lambda, \Lambda+1, \ldots$ Finally, $K+S = J$ is the total momentum taking into account spin. The corresponding quantum number is $J = K+S, K+S-1, \ldots, (K-S)$. Case *b* always occurs for states Σ, in which $\Lambda=0$ and S is not coupled with the interatomic axis. States Π and Δ often correspond to an intermediate state between *a* and *b*: in cases of weak rotation, *a*-type coupling is most frequently observed, and in cases of strong rotation, *b*-type coupling is observed. The molecules CN, N_2, etc. are good examples of this. There is obviously no distinction between modes *a* and *b* for singlet states, when $S = 0$, $\Omega = \Lambda$, and $K = J$. Of particular interest is the intermediate case when L and N are coupled. For all states with $\Lambda \neq 0$, this coupling splits every state with a definite J into two very close components (Λ-doubling).

Physically, this is due to the fact that all states with $\Lambda \neq 0$ are doubly degenerate (sign reversal of the angular momentum about the axis does not change the energy). $L-N$ coupling removes this degeneracy.

TABLE 4

Molecule	System (electron transition)	Vibrational transition	λλ in Å	Remarks
CO^+	$A^2\Pi_i - X^2\Sigma^+$ (system of cometary tails)	6 – 0	3416 – 3431	Doublet structure. Short-wave components slightly brighter than long-wave components. Low rotational temperature (~300°K)
		5 – 0	3580 – 3594	
		6 – 1	3695 – 3709	
		4 – 0	3781 – 3802	
		3 – 0	4001 – 4024	
		4 – 1	4124 – 4140	
		2 – 0	4251 – 4274	
		1 – 0	4544 – 4569	
	$B^2\Sigma^+ - A^2\Pi_i$ (Baldet-Johnson system)	2 – 1	3509	System intensity low, since: 1) transitions take place between excited electron levels; 2) excitation by $B^2X^+ - X^2\Sigma^+$ transitions occurs, in the far ultraviolet at λλ 2190 – 2042.
		1 – 0	3709 – 3726	
		0 – 1	4231	
N_2^+	$B^2\Sigma_u^+ - X^2\Sigma_g^+$ (negative system)	3 – 2	3545	Bands are usually weaker than CO^+ bands, and also wider; similar to N_2^+ bands in twilight-arc spectrum. Rotational temperature higher than for CO^+.
		1 – 0	3580	
		0 – 0	3913.7	
		1 – 2	4231	
		0 – 1	4274	
CO_2^+	$^2\Pi_u - {}^2\Pi_g$	1 – 0	3378	Extends in the tail over lesser distances than CO^+ and N_2^+.
		2 – 1	3388	
		0 – 0	3509	
		2 – 2	3545	
		0 – 1	3674	
		1 – 2	3695	
		0 – 2	3839	
OH^+	$^3\Pi_i - X^3\Sigma^-$	0 – 0	3594 – 3616	In head and initial part of tail. Relatively weak intensity.
CH^+	$^1\Pi_i - {}^1\Sigma^+$	1 – 0	3954 – 3972	Same
		0 – 0	4225 – 4254	P, Q and R branches

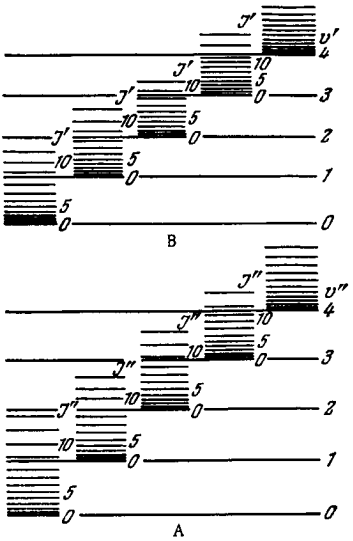


FIGURE 132

No degeneracy in this sense is observed for Σ states. However, we can distinguish between Σ states whose wave functions keep the same sign when reflected from a plane through the axis (Σ^+) and states whose wave functions reverse the sign during such a reflection (Σ^-). Transitions are permitted only between Σ states of like sign.

If the molecule is homonuclear (consists of two identical atoms), then it has a center of symmetry. A wave function may either conserve or change its sign when reflected at the center of symmetry. The states are classified correspondingly as even (g) and odd (u). Transitions are permitted only between states of different parity. When the nuclei are permutated, the wave function may either conserve or change its sign. Correspondingly, symmetric (s) and antisymmetric (a) states can exist, and transitions are permitted only between states of the same symmetry. At the rotational levels, the s and a states alternate (even J corresponds to s-symmetry and odd J to a-symmetry, or vice versa). Each electron state is thus split into a series of sublevels differing in their rotational energy, as determined by the quantum numbers J or K.

The vibrational motion of the molecule, which is accompanied by a variation in the internuclear distance r, must also be taken into consideration. The pattern of rotational sublevels is repeated for each vibrational state characterized by a definite vibrational quantum number v ($v = 0, 1, 2, \ldots$). Figure 132 is a diagram of the vibrational-rotational levels for two electron states A and B.

When a molecule has several different electron states with equal multiplicities and equal Λ, these states are differentiated by Latin letters preceding the term symbol (for example, $B^3\Sigma_u^+$). The ground-state level is generally designated by the letter X.

Transitions may occur between different electron states (electron transitions) and also between the sublevels of a single electron state (vibrational, rotational, and vibrational-rotational transitions). In cometary spectroscopy, electron transitions are of major significance, and each such transition is accompanied by simultaneous changes in both the vibrational and the rotational energy, so that here rotational-vibrational electron transitions occur. The frequencies emitted during these transitions are defined as follows:

$$\nu = \frac{\Delta E}{hc} = \frac{1}{hc}(E'_{el} - E''_{el} + E'_{vib} - E''_{vib} + E'_{ro} - E''_{ro}) =$$

$$= \nu_{el} + \nu_{vib} + \nu_{ro} = \nu_{el} + \omega'_e\left(v' + \frac{1}{2}\right) - \omega'_e x'_e\left(v' + \frac{1}{2}\right)^2 + \ldots -$$

$$- \omega''_e\left(v'' + \frac{1}{2}\right) + \omega''_e x''_e\left(v'' + \frac{1}{2}\right)^2 - \ldots + B_v J(J'+1) -$$

$$- D'_v J'^2 (J'+1)^2 + \ldots - B''_v J''(J''+1) + D''_v J''^2 (J''+1)^2 - \ldots, \qquad (16)$$

where ω_e, x_e, B_v, and D_v are constants, those with a single prime referring to the upper state and those with a double prime to the lower state. Here, the selection rules $\Delta J = 0, \pm 1$, and for b coupling $\Delta K = \pm 1$ as well (for $\Sigma - \Sigma$ transitions) or $\Delta K = 0, \pm 1$ (for the rest of the transitions) are observed: since the molecular vibrations are anharmonic, no selection rules are imposed upon Δv. To each pair of values of v' and v'', there corresponds a certain band with a more or less developed rotational structure. The transitions with $\Delta J = -1, 0, +1$, respectively, form the branches P, Q, and R. With b coupling, the strongest transitions are those with $\Delta K = \Delta J$. Thus, for example, for the violet CN^2 ($^2\Sigma - {}^2\Sigma$) system in cometary spectra only the R and P branches are pronounced.

Other conditions being equal, the intensity I_{ro} of a rotational transition is determined by the population of the level from which the transition occurs. As a first approximation, this intensity is

$$I_{rc} \sim (2J+1)\, e^{-BJ(J+1)\, hc/kT}, \qquad (17)$$

where $2J + 1$ is the statistical weight of the state; the exponential function is Boltzmann's factor. This expression is a maximum when

$$J = \sqrt{\frac{kT}{2Bhc}} - \frac{1}{2}, \qquad (18)$$

and we can thus determine the equivalent effective rotational temperature T_{ro} from the shape of the profile of the rotational branch. The properties of molecular spectra are discussed in greater detail in texts such as that of Kondrat'ev.

By analyzing the intensity distributions in the bands, we can make some important conclusions concerning the emission mechanism for cometary gases. In their study of the distribution of total emission along the tail of Halley's comet (1910 II), K.Schwarzschild and Kron expressed the opinion that the emission of cometary gases is due to a resonance re-emission of sunlight. The quantitative calculations of Zanstra and Wurm verified the plausibility of this explanation. However, the decisive proof of the resonance-fluorescent nature of cometary emission is provided by spectrophotometric analyses of comets; this was first demonstrated by Swings for the violet CN band (λ 3883). Actually, if CN is excited only due to the absorption of solar radiation by molecules occupying the lowest vibrational level

of the electron ground state, then the population of the excited level K' will be

$$N_{K'} \sim N''_{K'+1} p_P J_P + N''_{K'-1} p_R J_R. \tag{19}$$

Here, $N''_{K'+1}$ and $N''_{K'-1}$ are the populations of the rotational levels $K'' = K' + 1$ and $K'' = K' - 1$ of the electronic and rotational ground states from which electrons may pass to the level K'; p_P and p_R are the corresponding transition probabilities; and J_P and J_R are the intensities of solar radiation at the wavelengths corresponding to P transitions ($K'' = K' + 1$) and R transitions ($K'' = K' - 1$), as corrected for the Doppler shift caused by the radial velocity v_r of the comet relative to the sun.

For any distribution of $N''_{K''}$ the population $N'_{K'}$ will depend on the distribution of energy in the solar spectrum. In particular, the strong Fraunhofer lines of FeI (λ 3878.02 and λ 3878.57) should lead to anomalously low intensities of the P lines and R lines corresponding to K' values between 3 and 8 (depending on the Doppler shift), and for any v_r values $P(6)$ and $P(7)$ will always lie inside the minima. All these properties are actually observed. A quantitative test of Swings's suggestion was first carried out by McKellar for the CN and CH bands of some comets. A Boltzmann distribution was assumed for $N''_{K''}$, and the rotational temperature was chosen so as to achieve the best fit between the observed profiles and those calculated from (19). Here, all the observed intensity maxima corresponded to maxima of the calculated profile. The theoretical profile calculated from (17) neglecting the solar spectrum is smooth, without any secondary maxima.

A fluorescent excitation mechanism was also verified for most of the other molecular bands of OH, NH, CH^+, and C_2. An especially elegant confirmation of the fluorescent mechanism was obtained recently for the spectrum of Comet Mrkos 1957 d (1957 V) (Figures 133 and 134). Figure 133 shows the rotational structure of the CN band (0 — 0), calculated assuming a resonance-fluorescent excitation mechanism and taking into account the solar Fraunhofer lines and their Doppler shifts. When the Fraunhofer lines are neglected, the rotational maxima reach up to the dashed curve. Figure 134 shows the observed structure of this same band. A comparison of the two figures reveals a complete correspondence between the theoretical and experimental curves. So far, no bands of the cometary spectra have been found whose fine structures cannot be explained on the basis of the resonance-fluorescent excitation mechanism.

The fluorescent mechanism also explains satisfactorily the observed rotational temperature. The absorption in the R branch is usually stronger than in the P branch, so that the population of the level $K' = K + 1$ is higher than that of $K' = K - 1$ in the excited state. In the emission from $K' = K + 1$, the transition $P(K+2)$ is stronger than $P(K)$, so that the upper K'' levels become overpopulated. Therefore, a high-temperature distribution of electrons over the rotational levels may be created, especially when the Q branch is weak. Whether this distribution actually occurs, depends on the degree of stability of the distribution, that is, on the probability of rotational transitions to levels with lower K. For heteronuclear molecules, such transitions are permitted, so that a low-temperature distribution is observed (CH, OH, NH, CN, and CH^+). In homonuclear molecules, such as $C^{12}C^{12}$, purely rotational transitions are forbidden,* and a high-temperature

* Actually, in a purely rotational transition $\Delta J = \pm 1$, which is forbidden due to differences in the symmetries (a, s) of adjacent rotational levels.

distribution is observed ($T \approx 3000$—4000 °K). The situation is analogous for the vibrational temperatures: these are either low (in heteronuclear molecules) or high (in homonuclear molecules), depending on the molecular symmetry.

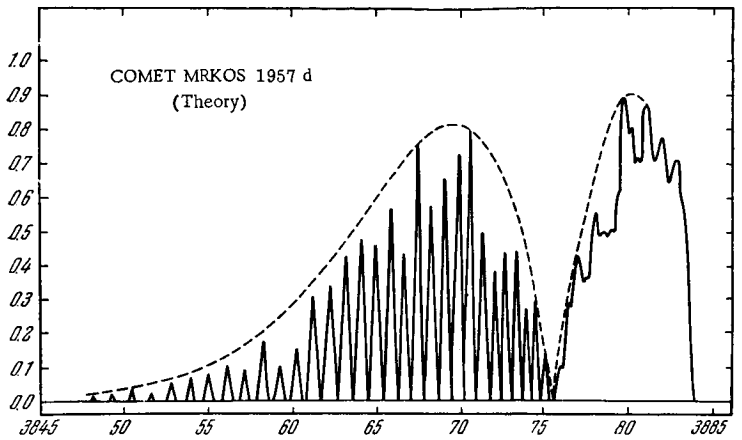


FIGURE 133

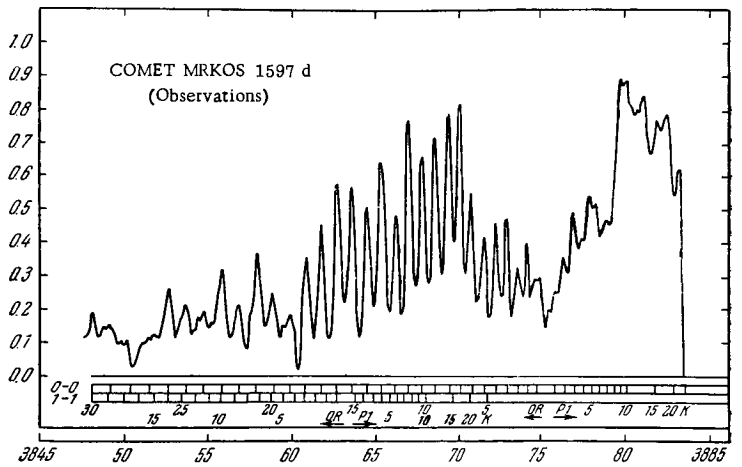


FIGURE 134

Since for resonance-fluorescence the decisive factor determining the rotational temperature T is the ratio of the mean interval between successive absorptions to the mean lifetime of the excited rotational states, therefore this temperature can be deduced by observing the cometary spectra.

For instance, Hunaerts, in his analysis of the profile of the OH band $(0-0)$, found that it could be adequately interpreted only if we set $K'' = 0$ (it could not be interpreted assuming a Boltzmann distribution over K''). Hence it follows that the probability of purely rotational transitions for OH is greater than $\frac{1}{\tau} = 2 \cdot 10^{-3}$ sec^{-1} (here τ is the time between two successive absorptions at a distance of 1 A.U. from the sun; it is determined from

the known (0—0) transition probability for OH and from the intensity of solar radiation).

Swings's fluorescent mechanism also explains the so-called Greenstein effect. If the image of the nucleus is projected onto the middle of the spectrograph slit, then in the spectra of CN and CH there may be differences of intensity between the "top" and the "bottom" of a given rotation line. For different lines these intraline variations may be expressed differently, and they may even have opposite signs. Hence it follows that the relative intensities of the rotational lines of a single band will differ in different parts of the head. This variation of the relative intensities is explained satisfactorily provided we assume differential radial (relative to the sun) velocities of around 1 to 3 km/sec within the head. As a typical example, let us consider the intensity distribution (expressed in arbitrary units) in the head of Comet Mrkos (1957 V), in the light of the $R(9)$ and $R(17)$ lines of the CN band (0—0) (Figure 135). The distribution obtained by Greenstein during observation was sharply asymmetric: in the $R(9)$ line the western (sunward) side was brighter, while in the $R(17)$ line the eastern side (that facing the tail) was more intense. If the brightness asymmetry could be attributed to the actual distribution of matter in the head or to certain instrumental effects, then it would remain constant for both lines. The effect is explained by the fact that the $R(9)$ and $R(17)$ lines lie in the wings of strong Fraunhofer lines of the solar spectrum (Figures 136 and 137). Due to the differential velocity distribution in the head, the western part of the head was displaced in the direction of higher λ, and the eastern part was shifted toward lower λ relative to the head center. Of course, the displacement was over the same interval for both lines, but in the case of $R(9)$ it enhanced the western part and weakened the eastern part relative to the center, while for $R(17)$ the opposite occurred. The origin and nature of the differential motions observed spectroscopically in the head have still not been explained properly. Hunaerts suggests that these motions are manifestations of the rotation of the nucleus; Swings and his co-workers, on the other hand, attribute them to the radial ejection of matter from the nucleus.

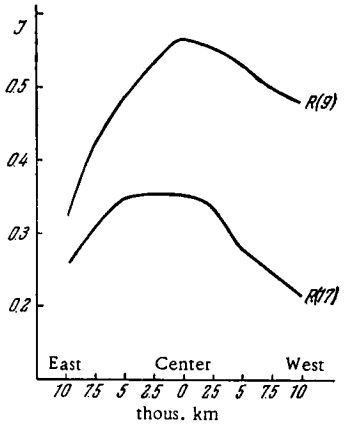


FIGURE 135

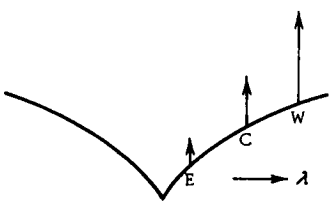


FIGURE 136

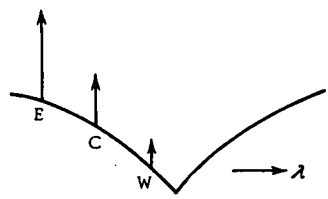


FIGURE 137

Most of the molecules in cometary atmospheres are free radicals. These may exist directly in the solid nucleus of the comet at very low temperatures or they may form due to the dissociation and ionization of some primary, or "parent," molecules. Some probable "parent" molecules are CH_4, NH_3, $(CN)_2$, and certain other more or less complex organic and inorganic molecules.

A very up-to-date review of cometary spectroscopy is given in the "Atlas of Representative Cometary Spectra," by Swings and Haser.

Most cometary spectra possess in common certain properties of variation with heliocentric distance. In particular, the various emissions generally appear in the order given in Table 5.

TABLE 5

r, A.U.	Nature of spectrum
>3	Sunlight reflected by dust particles or scattered by polyatomic molecules.
3	The (0 — 0) CN band appears; its fairly high abundance in the comet head is immediately apparent.
2	Emission bands of the triatomic molecules C_3 and NH_2 appear; these are much shorter than the CN lines; the C_3 intensity is higher than the intensity of the (0 — 0) CN band.
1,8	Swan bands appear; their normal abundance in the head, as well as their high rotational and vibrational temperatures, are immediately apparent.
1,5	All the other molecular emission bands of the head are observed (OH, NH, CH). Starting here, and at smaller r, the tail emissions appear: CO^+, N_2^+, and CO_2^+ have approximately the same distributions, and the CH^+ and OH^+ emissions are less extended.
$<1,5$	The band intensities increase to such an extent that their structures often become apparent. The profiles of the homonuclear molecules C_2 and C_3, which do not depend much on r, remain almost unchanged with further decrease of r. The profiles of heteronuclear molecules (CN, CH, OH, NH) have more developed rotational structures as r decreases. In particular, the (0 — 0) CN band, which when first observed has rotational lines up to $J = 10$, has many more rotational lines (J up to several tens) when r is small.

The relative intensities of the various emissions differ for different comets and they may fluctuate irregularly in time. Some comets have only a continuous spectrum and the (0—0) CN band, or they may have even only a continuum at all r (for example, 1910 I). Most comets, however, have certain regularities in common, as far as the distribution of the relative intensities of the various emissions is concerned.

1. Continuous spectrum. This is always predominant at large r; the strong continuum observed in the range of r from 1.5 to 1 A.U. generally remains for smaller r also.

2. Molecular emissions. For the blue-violet bands of C_3 (λ 4052), CN (the 0—1 band near λ 4193), CH (λ 4313), and C_2 (the 2—0 band near λ 4381) in the central part of the head, the relative intensities C_3/CN, CN/CH, and CN/C_2 have a tendency to increase on the whole with increasing r.

The relative intensity $\frac{OH}{NH}$ in the ultraviolet also increases with r in some comets. The rotational and vibrational temperatures of both heteronuclear and homonuclear molecules increase as r decreases.

§97. **The apparent brightness of a comet. Masses and densities of cometary atmospheres.** For a molecular spectrum, the brightness of a comet is determined by the number of molecules N in the ground electron state, since the lifetimes of the excited electron states are very short. If we agree to consider only one chemical species of molecule, then the number of photons scattered by the atmosphere per unit time will be

$$\sum_{v'v''} N p_{v'v''} J_{v'v''}, \tag{20}$$

where $p_{v'v''}$ is the probability of a vibrational electron transition between the vibration levels v' and v'', and $J_{v'v''}$ is the intensity of the solar radiation in the corresponding frequency range.

Next, by introducing the mean frequency $\nu_{v'v''}$ of the particular band system, we can write the energy emitted by the atmosphere per unit time as

$$E = N \sum_{v'v''} p_{v'v''} J_{v'v''} h\nu_{v'v''}. \tag{21}$$

The luminous flux Φ emitted by the comet is

$$\Phi = \frac{N}{M} \sum_{v'v''} p_{v'v''} J_{v'v''} h\nu_{v'v''} V_{v'v''}, \tag{22}$$

where $M = 1.61 \cdot 10^6$ erg/sec·lumen is the mechanical equivalent of light, and $V_{v'v''}$ is the visibility coefficient for the frequency $\nu_{v'v''}$. The illumination produced by the comet in the neighborhood of the earth is

$$\frac{\Phi}{4\pi\Omega^2} \text{ phots.}$$

Since the illumination produced by a star of zero magnitude is $2 \cdot 10^{-10}$ phot, the brightness of the comet can be expressed in stellar magnitudes as

$$m = -2.5 \lg \left(\frac{\Phi}{8\pi\Omega^2} \cdot 10^{10} \right). \tag{23}$$

If the cometary brightness is produced mainly by Swan bands, which lie near the sensitivity maximum for the eye (λ = 5550 Å), then Φ can be represented approximately as

$$\Phi = \frac{N}{M} B_{1-2} J_\nu h\nu, \tag{24}$$

where $\nu = \frac{c}{\lambda} = 5.4 \cdot 10^{14}$, and B_{1-2} is the Einstein coefficient for the C_2 electron transition $A^3\Pi_g \rightarrow X^3\Pi_u$. Alternatively, in terms of the oscillator strength

$$f_{1-2} = \frac{ch\nu m_e B_{1-2}}{4\pi^2 e^2}, \tag{25}$$

which according to Clementi is $\sim$0.05, we can write the flux as

$$\Phi = \frac{N}{M} \frac{4\pi^2 e^2}{cm_e} J_\nu f_{1-2} = \frac{1}{5} \frac{\pi^2 e^2}{cm_e} J_\nu \frac{N}{M}, \tag{24'}$$

from which

$$m = 2.5 \lg \left[\frac{N}{4M} \frac{\pi e^2}{cm_e \Omega^2} J_\nu \cdot 10^9 \right]. \tag{26}$$

Moreover, since

$$J_\nu = W \frac{2\nu^2}{c^2} \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1}, \tag{27}$$

where

$$W = \frac{1}{2}\left(1 - \sqrt{1 - \frac{r_\odot^2}{r^2}}\right) \approx \frac{1}{4}\frac{r_\odot^2}{r^2} = 5.5 \cdot 10^{-6} r^{-2} \tag{28}$$

is the dilution coefficient, therefore if we assume $T = 5 \cdot 10^3$ °K to be the temperature of the solar photosphere, we will obtain

$$m_\rho = m - 5 \lg \varrho = 2.5 \lg 0.8 \cdot 10^{35} - 2.5 \lg N,$$

where ϱ is in astronomical units, or

$$N = 0.8 \cdot 10^{35 - 0.4 m_\rho}. \tag{29}$$

This value gives only the lower limit of N. In reality, cometary atmospheres generally contain close to a dozen other kinds of molecules, all roughly as abundant as C_2.

By applying an analogous formula, Poloskov estimated the total number of molecules in the heads of some of the brightest comets (Table 6).

TABLE 6

Comet	N	R, cm	n	n_0	λ, cm	λ_0, cm
1843 I	10^{34}	$5 \cdot 10^8$	$2 \cdot 10^7$	$1.6 \cdot 10^{12}$	10^8	10^3
1882 II	10^{38}	$5 \cdot 10^8$	$2 \cdot 10^{11}$	$1.6 \cdot 10^{16}$	10^4	10^{-1}
1910 I	10^{35}	$5 \cdot 10^9$	$2 \cdot 10^5$	$1.6 \cdot 10^{12}$	10^{10}	10^3
1910 II	10^{34}	$5 \cdot 10^9$	$2 \cdot 10^4$	$1.6 \cdot 10^{11}$	10^{11}	10^4

In the table, R is the head radius for which N is estimated, n is the mean number of molecules per cm^3 of head volume, and $n_0 = \frac{N}{4\pi R_0^2 R}$ is the number of molecules per cm^3 of head volume in the immediate proximity of the comet nucleus. Here it is assumed that the density of the atmosphere obeys an inverse-square law with respect to the distance from the center of the nucleus:

$$n = n_0 \frac{R_0^2}{R^2}, \tag{30}$$

where the nuclear radius R_0 is 10 km, and λ and λ_0 are the corresponding mean free paths of the molecules, calculated as $\frac{1}{\sigma n}$, where $\sigma = 10^{-15}$ cm^2 is the gas-kinetic cross section.

In the heads of the faintest comets, the gas density is even lower. For instance, in his photoelectric observations of the intensity of the $(1-0)\,C_2$ band, Oiseau found that for $f = 0.035$ the density of C_2 in the head of Comet Mrkos (1955 III) was $3 \cdot 10^3\,cm^{-3}$ at a distance of $5.3 \cdot 10^8$ cm from the nucleus and $67\,cm^{-3}$ at a distance of $5.3 \cdot 10^9$.

The density of matter in cometary heads, even comparatively bright ones, is thus fairly low, and intermolecular collisions are actually negligible

throughout the head volume, except in the region immediately adjoining the nucleus, where such collisions may have some importance. An exception to this rule is Comet 1882 II, which was also exceptional in many other respects because of its small q (10^{-2} A.U.).

The number N is apparently related to the rate of gas liberation from the cometary nucleus and to the mean lifetime τ of a molecule before it dissociates or ionizes. If N_0 is the number of "parent" molecules in the cometary atmosphere, and N is the number of "daughter" molecules producing the optically observed spectrum, then according to Wurm

$$\left.\begin{aligned}\frac{dN_0}{dt} &= Q - \alpha N,\\ \frac{dN}{dt} &= \alpha N - \beta N,\end{aligned}\right\} \tag{31}$$

where Q is the number of "parent" molecules introduced into the atmosphere per unit time, and α and β are the reciprocals of the lifetimes of the "parent" and "daughter" molecules. After integration, we obtain

$$\left.\begin{aligned}N_0 &= \frac{Q}{\alpha} - C_0 e^{-\alpha t},\\ N &= \frac{Q}{\beta} - C_1 e^{-\beta t} - \frac{\alpha C_0}{\alpha - \beta} e^{-\alpha t}.\end{aligned}\right\} \tag{32}$$

For t much greater than $\frac{1}{\alpha}$ and $\frac{1}{\beta}$, we have

$$\left.\begin{aligned}N_0 &= \frac{Q}{\alpha},\\ N &= \frac{Q}{\beta}.\end{aligned}\right\} \tag{33}$$

Consequently, under quasistationary conditions N is equal to the lifetime τ of the molecules multiplied by the rate of liberation of "parent" molecules from the nucleus; it is independent of the rate of breakdown of the "parent" molecules.

The densities of the dust atmosphere were calculated by Whitney and others. Whitney assumed the surface properties of the dust particles to be the same as those near the surface of Ceres. Then, since Ceres, which has a diameter of 770 km, has at opposition an absolute magnitude of $3^m.7$, Whitney obtained the following expression for the decimal algorithm of the total mass of dust particles:

$$\lg M = \lg \delta h + 2 \lg \left[\frac{r\varrho\, 385 \cdot 10^5}{\Phi^{1/2}(\alpha)}\right] - 0.40m + 2.10, \tag{34}$$

where m is the apparent magnitude of the comet, δ is the density of a dust particle, h is its radius, and Φ is the phase function (the ratio of the brightness at phase angle α to the brightness at opposition); r and ϱ are expressed in astronomical units, and all the other quantities are in CGS units.

As an example of a determination of the density of the dust material, we may quote Liller's result for the type-II tail of Comet Arend-Rowland near the maximum of its development; he obtained $M \approx 10^{11}$ g for $h = 0.3\mu$, which gives a mean distance of about 4 m between dust particles at a distance of $2.1 \cdot 10^6$ km from the nucleus. For the type-II tail of Comet Mrkos, the same author gives a mass of $\sim 2.10^{14}$ g for $h = 0.3\mu$ and a mean distance of 3 m between particles $4.4 \cdot 10^6$ km away from the nucleus. In the USSR, the

surface photometry of comets has also been developed by Rijves, Konopleva, and others; they have established that the volume density n actually does decrease as the square of the distance R from the nucleus.

Vorontsov-Vel'yaminov analyzed the brightnesses of monochromatic images of Comet 1943 I (Whipple-Fedtke-Tevzadze) taken with a prism objective. He found the total number of molecules in the head to be: $N(C_2) = 7 \cdot 10^{32}$, $N(CN) = 1 \cdot 10^{33}$, and $N(C_3) = 3 \cdot 10^{31}$; for the observed head radius of $3.7 \cdot 10^5$ km, these figures give a density of $\sim 1\ cm^{-3}$ at the periphery of the head (assuming a square-law density drop with the distance from the nucleus).

Theoretical isophotes of a cometary head were plotted by Mokhnach and others for various assumptions concerning the ejection of molecules from the surface of the nucleus. In particular, when the ejection from the surface of the nucleus is proportional to the solar illumination, the isophotes (according to Mokhnach) have the form of slightly deformed circles with their centers displaced sunward relative to the nucleus. This, according to Mokhnach, causes a systematic error in the determination of the nucleus position from observations.

§98. **Differentiation of matter in a cometary atmosphere under the influence of solar heat.** The observed differentiation of the matter in a comet is caused by two factors: 1) differences in the acceleration a (producing the three types of tails), and 2) differences in the mean lifetime τ of molecules before dissociation or ionization (causing the separation of neutral molecules in the head and of ionized molecules in type-I tails).

The dust particles are impelled by the radiation-pressure force F of the sun, which is equal to the volume density of solar radiation $u = \frac{q}{c}\left(\frac{r_\odot}{r}\right)^2$ (where q is the solar constant and c is the velocity of light) multiplied by the effective cross section σ of the particle. A spherical particle of radius h and density δ acquires the following acceleration under the action of the force F:

$$a = \frac{3}{4}\frac{u}{h\delta}Q, \tag{35}$$

where $Q = \sigma/\pi h^2$ is the ratio of the effective cross section to the geometrical section.

The acceleration in this case may be measured conveniently in units of the acceleration of the [solar] force of gravity $a_0 = \frac{k^2 M_\odot}{r^2}$. Then, the dimensionless relative acceleration is

$$1 + \mu = \frac{3q}{4c}\frac{r_\odot^2}{k^2 M_\odot}\frac{Q}{h\delta} = 0.57 \cdot 10^{-5}\frac{Q}{h\delta}\ (h \text{ in cm},\ \delta \text{ in g/cm}^3). \tag{36}$$

The section $Q = 1$ in the limiting case when $h \gg \lambda$ and the refractive index of the particle $n \to \infty$ (metallic, perfectly conducting particles, or dielectric, perfectly reflecting particles). Then, we have

$$1 + \mu = \frac{0.57 \cdot 10^{-5}}{h\delta}. \tag{37}$$

For comparable h and λ and for $n \neq \infty$, the quantity Q must be determined taking into consideration the absorption and scattering of light by the particle according to Mie's general theory. Levin has used Mie's theory to

calculate the values of $1+\mu$ for stone particles (real n) and iron particles (complex n of the form $a+bi$). These data are given in Table 7.

TABLE 7

$\frac{d}{2}$, mμ	$1+\mu$		
	$n=1.5$ $\delta=3$	$n=2$ $\delta=4$	$n=1.57\times(1-i)$ $\delta=6$
50	0.05	0.3	1.7
100	0.3	1.0	1.5
150	0.5	1.2	—
200	0.5	1.2	—
300	0.5	0.7	—
400	0.4	0.5	—

These accelerations correspond to the values of $1+\mu$ observed in the dust of type-III tails.

The accelerations in type-II tails can be explained in two ways. On the one hand, if we assume a particle density lower than that in Levin's table or if we assume that the particles are considerably unspherical, then we obtain correspondingly higher $1+\mu$. These assumptions can be justified by considering various terrestrial minerals, for example, hydrocarbon derivatives having $\delta \approx 1$ with n as high as 2.

On the other hand, light exerts appreciable pressure on the molecules as well. In 1909 Lebedev staged his famous experiments in which he proved that radiation pressure is exerted upon gas molecules, and he showed this pressure to be proportional to the selective absorption of light. The modern theory of radiation pressure in gases was proposed by Pauli on the basis of quantum-mechanical concepts and then generalized by Wurm. This theory is as follows. A molecule absorbing a photon of energy $h\nu$ acquires its momentum $\frac{h\nu}{c}$, which is directed away from the sun. When it emits this photon, the molecule recoils and acquires additional momentum. In the case of forced emission, the additional momentum compensates for the absorbed momentum, and the net result is zero. In the case of spontaneous emission, the additional momentum can be oriented arbitrarily, and the mean recoil will be zero. In comets virtually no forced emission occurs, and the force acting on the molecules is equal to the momentum absorbed per unit time. This momentum is $\frac{B_{12}J_\nu h\nu}{c}$, where J_ν is the intensity of solar radiation in the frequency interval from ν to $\nu+1$ at the distance of the comet from the sun, and

$$B_{12}=\frac{4\pi^2e^2}{m_e h\nu c}f_{12} \tag{38}$$

is the corresponding absorption coefficient (Einstein coefficient).

By substituting J_ν from (27) and (28) and dividing by the molecular mass m_m and the acceleration of the sun's attraction $a_0\left(\frac{r_\odot}{r}\right)^2$, we now obtain the final expression for the relative acceleration:

$$1+\mu = C\frac{f_{12}}{m_m}\frac{1}{e^{\frac{h\nu}{kT}}-1}, \tag{39}$$

where

$$C = \frac{2\pi^2 e^2 h\nu^3}{m_e a_0 c^4} \frac{r_\odot^2}{r_\delta^2}.$$

The characteristic wavelengths responsible for the selective radiation pressure are known from spectroscopic observations. The Einstein coefficients have been determined with sufficient accuracy only for a few molecules, but approximate values have been obtained from quantum-mechanical calculations. If we set $f_{12} \approx 0.05$, then we find $1 = \mu \approx 3$ for C_2 molecules, which is consistent with the corresponding accelerations observed in type-II tails.

In type-I tails, however, much higher accelerations are observed, corresponding to forces 20 to 200 times (and in exceptional cases even a thousand times) stronger than the force of solar attraction. Moreover, the radiation pressure is always directed radially away from the sun, while the gaseous tails can also have transverse motions in them. Thus, besides radiation pressure, we must seek some other reasonable explanation for the accelerations of the gaseous tails (see §99).

The lifetime τ of a molecule in a field of Planck radiation from the sun was studied by Wurm, Ambartsumyan, Poloskov, and others. It is clear that

$$\tau = \frac{1}{B_{12} J_\nu} = \frac{m_e c^3}{8\pi^2 e^2 \nu^2 W} \left(e^{\frac{h\nu}{kT}} - 1\right), \tag{40}$$

where ν is the frequency of the photodissociating radiation. The numerical values obtained using (40) are uncertain, because the effective ν, on which τ depends exponentially, has not been definitely determined. Poloskov, together with Herzberg and Phillips, has assumed that C_2 will photodissociate for $\lambda = 2000$ Å, and thus he obtains a τ value close to a half-hour. Ambartsumyan takes $h\nu = 15$ eV and obtains a correspondingly higher higher τ.

Nevertheless, it seems that the observed abundances of neutral molecules in cometary heads (the sequence NH_2, C_3, OH, NH, C_2, CN) are directly related to the mean molecule lifetimes in the field of Planck radiation from the sun, since molecules of relatively complex structure, which dissociate more readily, tend to concentrate near the nucleus.

§99. Type-I tails. The structures of ionized tails are determined, according to present-day conceptions (Alfvén and others), mainly by the effects which corpuscular streams from the sun have upon comets. When the head of a comet enters into a corpuscular stream, the ions produced in the head form in the "ready-made" magnetic field $\boldsymbol{H}$ of the stream, regardless of the ionization mechanism.

The motion of the ions is described by Maxwell's equations:

$$\operatorname{curl} \boldsymbol{H} = \frac{1}{c}\left(4\pi \boldsymbol{i} + \frac{\partial \boldsymbol{D}}{\partial t}\right) \approx \frac{4\pi \boldsymbol{i}}{c}, \tag{41}$$

$$\operatorname{curl} \boldsymbol{E} = -\frac{1}{c}\frac{\partial \boldsymbol{B}}{\partial t}, \tag{42}$$

where

$$\boldsymbol{B} = \mu \boldsymbol{H}, \quad \boldsymbol{D} = \varepsilon \boldsymbol{E}, \tag{43}$$

$$\operatorname{div} \boldsymbol{H} = 0, \tag{44}$$

$$\boldsymbol{i} = \sigma\left\{\boldsymbol{E} + \frac{1}{c}[\boldsymbol{v}\boldsymbol{B}]\right\}, \tag{45}$$

and by Ampére's law:

$$\varrho \frac{d\boldsymbol{v}}{dt} = \frac{1}{c}[\boldsymbol{i}\boldsymbol{B}] - \nabla p \tag{46}$$

(where p is the pressure, and ϱ is the density). From (41) and (45) we have

$$\operatorname{curl}\operatorname{curl} \boldsymbol{H} = \frac{4\pi\sigma}{c} \operatorname{curl} \boldsymbol{E} + \frac{4\pi\sigma\mu}{c^2} \operatorname{curl} [v\boldsymbol{H}]. \tag{47}$$

But since

$$\operatorname{curl}\operatorname{curl} = \operatorname{grad}\operatorname{div} - \Delta, \tag{48}$$

therefore, applying (44) and (42), we obtain from (47)

$$\frac{\partial \boldsymbol{H}}{\partial t} - \operatorname{curl} [v\boldsymbol{H}] = \lambda \Delta \boldsymbol{H}. \tag{49}$$

If the coordinate system moves with the medium, then $v = 0$ and

$$\frac{\partial \boldsymbol{H}}{\partial t} = D\Delta \boldsymbol{H}, \tag{50}$$

where $D = \frac{c^2}{4\pi\sigma\mu}$ can be interpreted as the coefficient of diffusion of the force lines.

In cometary heads, the electrical conductivity $\sigma = \frac{e^2 n\tau}{2m} \to \infty$, since the time τ between two successive electron collisions is very high. Consequently, $D \to 0$ and the force lines virtually do not diffuse. This also applies to the corpuscular stream. The force lines thus appear to be "frozen in" both in the corpuscular stream and in the head of the comet. The formation of this common magnetic field enveloping the comet and the corpuscular stream is depicted, according to Alfvén, in Figure 138.

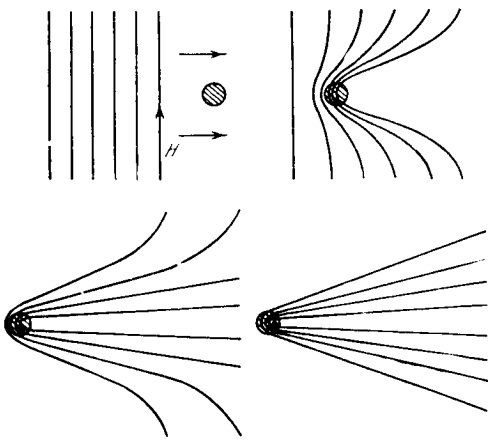


FIGURE 138

The cometary ions in the head move collectively relative to the stream, transporting the field with them, and gradually drifting along the force lines they form a ray system analogous to a system of force lines.

The rays, in full agreement with the observations, are rectilinear, gradually converging toward the axis of symmetry. The ray structure is

observed for a long time after the stream has passed by. The wave motions in the rays are explained in terms of Alfvén magnetohydrodynamic waves. Their formation has been illustrated by Alfvén for the simplest case of a homogeneous incompressible fluid.

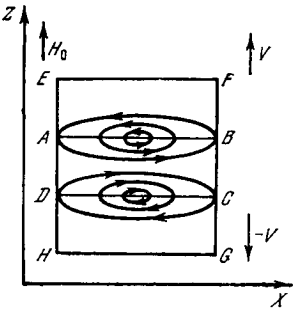


FIGURE 139

Let us consider an infinite conducting fluid, initially at rest except for a rectangular prism extending to infinity along the y axis and defining the rectangle $ABCD$ in the xz plane (Figure 139). This prism moves along the y axis with a velocity v. If a uniform magnetic field $\boldsymbol{H}_0$ pointing along the z axis is applied to the fluid, then an electric field $\boldsymbol{E} = \frac{\mu}{c}[v\boldsymbol{H}_0]$ pointing along the x axis will be induced in the prism. The field sets up a current i, which is shorted through the surrounding stationary fluid. A force $\boldsymbol{F} = \frac{\mu}{c}[\boldsymbol{i}, \boldsymbol{H}_0]$ pointing in the negative-y direction acts on the current i. This force slows down the moving parallelepiped and accelerates the initially stationary fluid in the adjoining volumes $ABFE$ and $DCGH$. This process continues, and as a result a magnetic wave is generated which transmits the state of motion along the magnetic field. Let us now calculate the propagation velocity of the wave.

If $\boldsymbol{h}$ is the field set up by the current $\boldsymbol{i}$, then the total field is $\boldsymbol{H} = \boldsymbol{H}_0 + \boldsymbol{h}$. We assume that all the vectors depend on z and t only. Then, from (41) and (42), we have $v_z = 0$, and since the displacement current is negligible, we have $\operatorname{div} v = 0$ and $i_z = 0$. Also, by choosing the proper y axis, we can make $i_y = 0$. Then

$$i_x = -\frac{c}{4\pi}\frac{\partial h_y}{\partial z}, \quad h_x = \text{const} = 0 \tag{51}$$

and from (46)

$$\frac{\partial v_y}{\partial t} = \frac{\mu H_0}{4\pi\varrho}\frac{\partial h_y}{\partial z}. \tag{52}$$

On the basis of (42), we have

$$\frac{\partial h_y}{\partial t} = -\frac{c}{\mu}\frac{\partial E_x}{\partial z},$$

and then from (45)

$$E_x = \frac{i_x}{\sigma} - \mu\frac{v_y}{c}H_0.$$

Hence,

$$\frac{\partial^2 h_y}{\partial t^2} = H_0 \frac{\partial^2 v_y}{\partial t\, \partial z} - \frac{c}{\sigma\mu} \frac{\partial^2 i_x}{\partial t\, \partial z}$$

or, by inserting (51) and (52), we obtain

$$\frac{\partial^2 h_y}{\partial t^2} = \frac{\mu H_0^2}{4\pi\varrho} \frac{\partial^2 h_y}{\partial z^2} + D \frac{\partial^3 h_y}{\partial z^2\, \partial t}. \tag{53}$$

For $\sigma \to \infty$, we have the wave equation

$$\frac{\partial^2 h_y}{\partial t^2} = V^2 \frac{\partial^2 h_y}{\partial z^2}, \tag{54}$$

where

$$V = \pm \sqrt{\frac{\mu H_0^2}{4\pi\varrho}} \tag{55}$$

is the propagation velocity of the magnetohydrodynamic wave, which depends neither on the frequency nor on the amplitude of the oscillations. For finite σ, the solution of the wave equation (53) can be sought in the form

$$h = h_0 e^{\alpha z + i\omega t},$$
$$v = v_0 e^{\alpha z + i\omega t}.$$

Here i is the imaginary unit, and ω, h_0, v_0, and α are constants, each of which, except for ω, may be complex. In order to find these constants, we can construct from (53) the equation

$$\omega^2 + \left(V^2 + i\frac{c^2\omega}{4\pi\mu\sigma}\right)\alpha^2 = 0$$

or

$$\alpha = \pm \frac{i\omega}{V}\left(1 + \frac{i\omega D}{V^2}\right)^{-\frac{1}{2}},$$

which, for weak damping, when $i\omega D \ll V^2$, gives

$$\alpha \approx \pm \left(i\frac{\omega}{V} + \frac{\omega^2 D}{2V^3}\right). \tag{56}$$

Here, the real part of α gives the reciprocal of the distance at which the amplitude drops to $1/e$, while the imaginary part shows that there is a phase difference between the velocity V and the excitation field h.

In cometary tails, σ is very high, $\mu = 1$, and the propagation velocity of the magnetohydrodynamic wave is

$$V = \sqrt{\frac{H_0^2}{4\pi\varrho}}. \tag{57}$$

This in principle enables us to find the cometary magnetic fields from the apparent velocity of the waves in the tails and from the tail brightness. In any case, (57) gives the correct orders of magnitude. For instance, for $H_0 = 10^{-5}$ oersted and $\varrho = 5 \cdot 10^{-21}$ g/cm^3 ($n = 10^2$ particles /cm^3), we have

$$V = 125 \text{ km/sec}.$$

On the whole, the ray system of the cometary tail should lie approximately in the plane through the comet and the corpuscular stream. Consequently, an observer in this plane will see a narrower tail than observers who view

the comet at other angles (provided the comet is not acted upon simultaneously by several differently oriented corpuscular streams). This conclusion can be verified by observation.

It is known, however, that the cometary ray systems are not particularly stable. A ray system may literally decay in a few hours or in some tens of hours, and may turn into a random intertwining of various elements and shapeless formations. Of course, it may become reconstructed just as rapidly.

The instability of the ray systems also has a logical explanation. The stability criterion for a corpuscular stream relative to the turbulence is given by the inequality

$$\left(1+\frac{\varrho_m}{\varrho_s}\right)\left(\frac{H_m^2}{4\pi}+\frac{H_s^2}{4\pi}\right)\geqslant \varrho_m v_s^2, \tag{58}$$

where the subscripts "m" and "s" refer to the interplanetary medium and the stream, respectively. If we assume $\varrho_s = 10^{-22}$ g/cm^3, $H_m = 0$ and $v_s = 10^8$ cm/sec, we obtain the following condition for the absence of turbulence:

$$H_s \geqslant 3\cdot 10^{-3} \text{ oe}, \tag{59}$$

or, for $v_s = 3\cdot 10^7$ cm/sec,

$$H_s \geqslant 3\cdot 10^{-4} \text{ oe}. \tag{60}$$

A corpuscular stream may thus easily become turbulent, making an ordered ray structure impossible.

The time required for the stream to become turbulent is

$$T_{\text{turb}} \sim \frac{l}{\sqrt{v_s^2-\frac{H_s^2}{4\pi\varrho_m}}},$$

where l is the width of the stream. If for $r\sim 1$ A.U., we have $l\sim 10^{13}$ cm, $H_s = 2\cdot 10^{-4}$ oe, $\varrho_m = 10^{-22}$ g/cm^3, and $v_s = 10^8$ cm/sec, then $T_{\text{turb}}\sim 10^5$ sec. Consequently, at a distance of $r=1$ A.U., part of the stream will still have ordered magnetic fields in it; however, as r increases, the percentage of ordered fields relative to the turbulent areas will drop considerably.

Interactions with corpuscular streams can explain not only the shapes of type-I tails but also the high accelerations of the cloud formations in the tails, and even the ionization process by which CO^+ ions are produced in comets.

The reasons for the high accelerations a have been considered by Poloskov, Biermann, Dobrovol'skii, and others. Poloskov calculated the upper limit for the acceleration due to radiation pressure, taking $f=1$ as the oscillator strength; however, since this limit is only about 50 times the acceleration of the solar force of gravity, radiation pressure alone cannot account for the high accelerations. Biermann explains the high a as being due to internal friction between the tail plasma and the plasma of a corpuscular stream. He obtained the following expression for a, if the electron temperature of the stream is $T_e = 10^4$ deg:

$$a = 10^{-4,3} n_\odot v_\odot \text{ cm/sec}^2,$$

where $n_\odot$ and $v_\odot$ are respectively the number of particles per cm^3 of volume and the velocity of the corpuscular stream in cm/sec. Dobrovol'skii has

pointed out the importance of the collective particle motions which arise when the tail interacts with a stream. It has been found recently that the collective motions become particularly important when there is a magnetic field "frozen into" the plasma of the stream or the comet. If, for example, a field H is "frozen into" an ionized cloud in a cometary tail, then charged particles cannot penetrate into the cloud further than the Larmor radius r_L, which is defined by setting the centrifugal force mv/r_L equal to the Lorentz force:

$$\frac{e}{c}[vH] = \frac{e}{c} vH_{\perp},$$

where $H_\perp$ is the component of H perpendicular to v. The particles can thus reach a depth of $r_L = \frac{mvc}{eH_\perp}$, or, for a solar corpuscular stream, a depth of $r_{eff} = \sqrt{r_e r_p}$, which is the geometrical mean of the Larmor radii for an electron and a proton. For $v = 10^8$ cm/sec and $H_\perp = 10^{-5}$ oersted, we obtain $r_{eff} \approx 10^7$ cm, which is much less than the characteristic dimension l of the cloud ($\sim 10^9$ cm). The stream therefore cannot penetrate into the cloud, and it appears to "flow around" it. The momentum transferred to the cloud per sec in this case is of the order of $n_\odot m_p v^2 \cdot \pi l^2$, and the acceleration is

$$a = \frac{dv}{dt} \approx -\frac{n_\odot}{n_k} \frac{m_p}{m_k} \frac{v^2}{l} = -\varkappa v^2.$$

Now, if we integrate with respect to time from 0 to t and with respect to velocity from the initial value $v_\odot$ to v, we obtain

$$v = \frac{v_\odot}{1 + v_\odot \varkappa t},$$

and the acceleration as a function of time is

$$a = -\frac{\varkappa v_\odot^2}{(1 + v_\odot \varkappa t)^2}.$$

The time required for a to decrease to $1/e$ of the initial value is

$$t_e = \frac{(\sqrt{e} - 1) v_\odot}{a_0} = \frac{\sqrt{e} - 1}{v_\odot \varkappa}.$$

Thus, the acceleration may vary over a wide range, from zero to virtually instantaneous entrainment of the cloud by a stream with a velocity $v_\odot$.

For instance, for Comet Morehouse (1908 II), the mean accelerations near the nucleus were 63 cm/sec^2, and they dropped to $1/e$ of this figure at distances of around $2 \cdot 10^{12}$ cm (distances which could be covered in about $5 \cdot 10^5$ sec). For $l = 10^9$ cm and $\frac{m_p}{m_k} = \frac{1}{28}$, this corresponds to $\frac{n_\odot}{n_k} = 10^{-4}$ and $v_\odot = 6 \cdot 10^7$ cm/sec, values which could have been expected. However, a rigorous theory describing the acceleration in a magnetic field has not yet been developed beyond the initial stages.

The ionization of cometary atmospheres was until quite recently the stumbling block of the physical theory of comets, since neither the ultraviolet radiation from the sun, nor X rays, nor impact ionization could ensure a sufficiently high rate of CO^+ production. The formation of an ionized atmosphere can be explained in a straightforward manner only in terms of a charge exchange between the stream protons and the neutral molecules of the cometary atmosphere:

$$H^+ + \begin{cases} CO = CO^+ + H \\ N_2 = N_2^+ + H \end{cases}$$

On account of the quasiresonant nature of the process (the ionization energies are: for H, 13.5 eV; for CO, 13.2 eV; and for N_2, 15.5 eV), the effective cross section σ is close to the gas-kinetic value ($6 \cdot 10^{-16}$ cm^2 for N_2 and 10^{-16} cm^2 for CO at velocities of 10^8 cm/sec).

The mean lifetime of a molecule before charge exchange is

$$\tau = (n_\odot v_\odot \sigma)^{-1}.$$

If $n_\odot = 10^2$ cm^{-3} and $v_\odot = 10^8$ cm/sec, then $\tau \approx 10^5$ sec. During this time, CO molecules can cover on the average a distance $s = v_0 t - \frac{at^2}{2}$, a distance which, for a velocity of ejection from the nucleus of v_0 $5 \cdot 10^4$ cm/sec and an acceleration of 0.5 cm/sec^2, is equal to $2.5 \cdot 10^9$ cm. This distance is comparable to the radius of the head of Comet Morehouse (10^{10} cm), which consisted almost entirely of CO^+.

The irregular fluctuations often observed in the intensity of the CO^+ ejection can also be explained quite naturally in terms of fluctuations in $n_\odot$. A comparatively high value of $n_\odot v_\odot$, for the corpuscular flux, 10 or 100 times the values measured with space rockets, is required to explain τ. Such a strong flux can be explained, according to Marochnik, by the cumulative effect of the magnetic field in the cometary head (see Figure 138), which enhances the value of $n_\odot v_\odot$ for the flux.

§100. The origin of comets. The first scientific theories on the origin of comets were proposed almost simultaneously by Laplace and Lagrange in 1813—1814. According to Laplace, comets are captured by the solar system from interstellar space; according to Lagrange, comets are permanent members of the solar system, originally ejected by the major planets during eruptions. Both theories were subsequently developed and still have supporters today. An especially popular theory is that of Oort and van Woerkom, according to which the comets form a large belt, similar to the asteroid belt, but lying far beyond the orbit of Pluto. Comets leave the belt and approach the sun due to the gravitational perturbations of stars. Oort and van Woerkom's theory was expanded by V. A. Krat within the framework of his new cosmogonic hypothesis. According to Krat, the solar system formed from a cloud of solid particles much like the nucleus of a comet. At the periphery of the cloud, beyond the present orbit of Neptune, conditions were unfavorable for the condensation of these particles into planets, so that the outer belt of minor bodies remained. According to the theory, comets emerge from the trans-Neptunian region not only due to the gravitational fields of stars but also on account of the mutual perturbations of the minor bodies and the gravitational perturbations of the major planets.

Future research in cosmogony will enable us to decide which of the theories explaining the origin of the comets is the closest to reality.

§101. Some unsolved problems in the physics of comets. 1. Cometary spectra in the far ultraviolet and in the infrared and radio ranges should be obtained. Extra-atmospheric stations will usually be necessary for this. Such spectra will enable us to establish the chemical compositions of cometary atmospheres, in particular, the nature of the "parent" molecules.

2. The nature of the dust particles in type-II tails should be determined. These have been studied by measuring the polarizations and the colors of

tails (relative to the sun) in the continuous spectrum and then comparing the results with the theory describing diffraction about dust particles of various sizes and compositions. This was studied in greatest detail by Liller, who decided that the particles were iron, with radii of 0.3 μ and with the ordinary density of iron. Other authors have maintained that the particles are stone and have low densities (up to 0.05 g/cm^3). Further detailed investigations are needed.

3. A quantitative theory of the ray systems and accelerations in ionized tails should be formulated.

4. The nature of the sudden ejections from cometary nuclei (halos) and their possible relations to solar activity should be studied.

5. Cometary nuclei should be simulated by releasing snowballs from rockets at high altitudes.

6. A rocket should be sent into a comet in order to study the composition of the head.

The Committee on Cometary Physics at the 11th Congress of the International Astronomical Union recommended that points 5 and 6 be carried out in the very near future, in particular with regard to the 1964 apparition of Encke's Comet.

§102. **General data on meteors.** A "meteor" is the name given to the combination of phenomena in the earth's atmosphere which accompanies the flight of a meteoric body. The kinetic energy of the meteoric body is used for heating, for exciting the visible glow of the meteor, and for ionization of the atmosphere. The thermal, luminous, and electrical energies of a meteor stand in approximately the following ratio to one another: $10^4 : 10^2 : 1$.

Meteoric bodies are classified as stream [shower] meteors and sporadic meteors. Stream meteors roughly follow the orbits of the comets producing them and they form meteor streams. All other meteoric bodies are classified as sporadic, and they show up as sporadic (occasional) meteors. These may be fragments of asteroids, remnants of meteor swarms which have dispersed or thinned out, etc. The pioneers of large-scale systematic meteor observations were Denning in England, Olivier in the USA, and Astapovich in the USSR.

The principal night-time meteor streams which form showers every year are listed in Table 8. The columns "Epoch" and "Radiant" give the dates of the beginning, maximum, and end of the shower and the Hoffmeister coordinates of the radiant for these epochs. For some showers data are given only for the epoch of maximum. It should be remembered that each shower has a certain radiation area. Therefore, the values of α and δ obtained by different observers may differ by as much as a few degrees.

The "Frequency" column gives the relative shower intensity according to Hoffmeister. The numbers in this column refer to the mean number of meteors visible in the entire sky with the naked eye for one hour during the epoch of maximum, when the radiant is at the zenith. They can be compared with the analogous hourly numbers of meteors not participating in showers (about 10 meteors per hour) which have daily and seasonal variations.

The "v_g" column gives the geocentric velocities of the meteors determined by radar methods (first three values in column) and photographically (last six in column). Finally, the last two columns give the average altitudes of the meteors, which obviously are directly related to v_g.

The principal day-time streams which have been studied using radar techniques are given in Table 9, following the general pattern of Table 8.

All these streams are identified with corresponding night-time streams, since their orbits meet the earth twice (also compare the night-time η-Aquarids and Orionids).

TABLE 8

Name	Radiant α	Radiant δ	Epoch	Frequency	v_g, km/sec	Parent comet	Mean altitude appearance, km	Mean altitude disappearance, km
Quadrantids	227°	+46	3 Jan	145	40	(Kozik-Peltier) 1939 I		
Lyrids	273	+35	12 Apr 22 Apr 24 Apr	40	48	1861 I		
Aquarids	335 338 345	−2 −1 0	29 Apr 5 May 21 May	120	64	Halley?		
Aquarids	332 343 347	13 17 17	25 July 3 Aug 10 Aug	40	43	?	100	90
Perseids	20 43 57	+46 +56 +59	20 July 11 Aug 19 Aug 11 Oct	300	60	1862 III	115	95
Orionids	94	+16	19 Oct 30 Oct	50	66	Halley?	120	100
Taurids	25 58 79	+16 +21 +20	24 Sept 13 Nov 10 Dec	25	30	Encke	100	78
Leonids	151	+21	16 Nov	20	71	1866 I	125	95
Geminids	110 113 117	+28 +30 +30	5 Dec 12 Dec 19 Dec	50	36	?		

TABLE 9

Name	Radiant α	Radiant δ	Epoch	v_g, km/sec	Parent comet	Corresponding night-time shower	Remarks
Arietids	39° 44 49	+18° +23 +29	30 May 8 June 18 June	38		Aquarids	The difference in the epochs and radiants of the showers related to Comet Encke is due to Jovian perturbations
ζ Perseids	51 61 69	+12 +22 +28	2 June 9 June 17 June	29	Encke	The small Arietid shower of 12 Oct	
β Taurids	85	+17	30 June	31	Encke	Taurids	

Table 10 lists some streams which are not observed annually. These may be either meteoric bodies moving in compact groups (and not dispersed over the entire orbit), which therefore cannot be observed every year but which when observed give heavy meteor showers (for example, the October Draconids), or else streams which due to orbit perturbation no longer meet the earth (the June Draconids).

TABLE 10

Name	Date	Years	Parent comet	Radiant	
				α	δ
Andromedids	27 Nov	1838, 1872, 1885, 1892, 1899	Biela	23°	+43°
June Draconids	28 June	1914—1921	Pons-Winnecke	211	+60
October Draconids	9 Oct	1926, 1933, 1946, 19....	Giacobini-Zinner	262	+52
Ursids	22 Dec	1945—1946	Tuttle	218	+79

Meteor brightness is measured in stellar magnitudes. The visual photometry of meteors was developed by Öpik, and their photographic photometry by Sytinskaya, Ivanikov, and others. The luminosity function for meteors is given in Table 11 according to the data of Cook, based on about 100,000 visual observations of sporadic and stream meteors.

TABLE 11

Stellar magnitude m, reduced to zenith	Relative number of meteors observed	Relative number of meteors, corrected for detectability of weak meteors
≤ 0	3	3
1	12	12
2	22	22
3	27	30
4	22	44
5	14	168

According to Levin, sporadic and stream meteors have different luminosity functions. For sporadic meteors, $\varkappa = \frac{A(m+1)}{A(m)} = 3$, and for stream meteors on the average $\varkappa = 2.3$, where $A(m)$ is the number of meteors in the interval $\left[m - \frac{1}{2}, m + \frac{1}{2}\right]$. In general, $A(m) = A(0)\,\varkappa^m$.

We can also find the mass distribution function $f(M)$ for meteors, which is the number of meteors per unit interval of the mass M, provided an additional assumption is made concerning the relation between the luminous intensity J of the meteor and its mass M. Generally it is assumed that $J \sim M^x$. Then, since $m = \frac{2.5}{y} \lg \frac{A}{A(0)} = -2.5 \lg \frac{J}{J(0)}$, where $y = 2.5 \lg \varkappa$, we find that

$A \sim M^{-s}$ and $f(M) = A\frac{dm}{dM} \sim \frac{1}{M^{1+s}}$, where $s = xy$. From this we can calculate the distribution function of the radii r of meteoric bodies:

$$f(r) = \frac{1}{r^{1+3s}}.$$

Next, by setting $x = 1$ and $y = 1$ ($\varkappa = 2.5$), we obtain the often-encountered expressions:

$$f(M) = M^{-2}, \quad \varphi(r) = r^{-4}. \tag{61}$$

Relations (61) have been confirmed telescopically and by means of radar observations (see Table 12, which gives the relative number of meteors entering the earth's atmosphere, according to Lovell).

TABLE 12

m	1	2	3	4	5	6	7	8	9
Number of meteors	2	2	6	28	53	143	292	790	2240

Measurements from space rockets and satellites, which have been summarized by Dabin and McCracken, give the following expression for the total particle flux J falling onto the earth (in $m^{-2}\,sec^{-1}$):

$$\lg J = -17.0 - 1.70 \lg M$$

for meteor masses in the range $10^{-10}\,g \leqslant M \leqslant 10^{-6}\,g$.

In order to calculate the total mass $\mathfrak{M}$ of meteoric matter falling onto the earth, these distribution functions must be extrapolated beyond the range of observed m to establish the numerical relation between m and M. This is impossible at present, however, without various simplifying assumptions. Different authors may thus disagree by several orders of magnitude in their estimates of $\mathfrak{M}$. A probable figure for the total mass is $\mathfrak{M} = 10^4$ tons/day.

§103. **Elements of the physical theory of meteors.** A meteoric phenomenon will be observed differently depending on whether the mean free path λ for the air molecules is larger or smaller than the radius r of the meteoric body. For $\lambda \gg r$, free molecular flow takes place. For $\lambda \approx r$, the surface of the meteoric body is screened by a cloud of reflected and vaporized molecules. The processes of heat and momentum transfer through this cloud become of decisive importance. Finally, for $\lambda \ll r$, the air flows around the meteoric body as a continuous fluid; a shock wave is formed.

Since observations show that the great majority of meteors ignite and burn out at altitudes of about 120—80 km, where the first two cases apply, the physical theory has been developed mainly for these conditions. The principal equations are as follows.

1. The braking equation is

$$M\frac{dv}{dt} = -\Gamma S \varrho v^2, \tag{62}$$

where S is the cross-sectional area of the meteoric body, ϱ is the

atmospheric density, and Γ is the resistance coefficient, which shows by what factor the momentum $S\varrho v^2$ acquired by the atmosphere per unit time (when the meteoric body captures all the molecules colliding with it) exceeds the actual resistance force acting on the meteoric body.

2. The vaporization equation is

$$\frac{dM}{dt} = -\Lambda \frac{S\varrho v^3}{2Q}, \tag{63}$$

where Q is the energy required to break down a unit mass of the meteoric body, and Λ is the coefficient of heat exchange, which shows what proportion of the relative kinetic energy $S\varrho v^3$ of the air molecules colliding with the body per unit time is used to break up the surface layer.

3. The luminosity equation is

$$\Phi = 4\pi J = -\tau \frac{dM}{dt} \frac{v^2}{2}. \tag{64}$$

where Φ is the luminous flux, J is the instantaneous luminous intensity of the meteor, and τ is the luminosity coefficient, showing what proportion of the kinetic energy of the particles separating from the meteoric body is used to excite the visible emission. The luminosity coefficient is generally assumed to vary approximately linearly with v:

$$\tau = \tau_0 v^p, \tag{65}$$

where $p \approx 1$.

The main object of the theory is to obtain a correct estimate of the coefficients Q and Γ, and particularly τ and V. The case $\lambda \gg r$ is comparatively simple: all the coefficients may be taken as constant for a particular meteor, although they depend on the nature of the meteoric body and on its velocity. On the average: $Q \approx 10^{11}$ erg/g; $\Lambda = 1$ for stony bodies and 0.75 for iron bodies; $\Gamma = 1.3$ for iron bodies, $\Gamma > 1$ for rapidly moving stony bodies, and $\Gamma = 1$ for slow stony bodies. For $\lambda \approx r$, the coefficients depend on v, ϱ, r, the shape of the body, and other factors (in particular, they depend on the distintegration mechanism: cathode sputtering, which prevails at high altitudes, gives way to a vigorous vaporization and blowing away of minute droplets at low altitudes). It follows from various experiments and theoretical calculations that the numerical values of the coefficient may lie approximately between the following limits: $1 > \Lambda > 10^{-2}$; $1.5 > \Gamma > 0.4$; $2 \cdot 10^{11}$ erg/g $> Q > 8 \cdot 10^{10}$ erg/g; $5 \cdot 10^{-7}$ lumen per erg/sec $> \tau > 5 \cdot 10^{-8}$ lumen per erg/sec (for $p = 1$).

Of all the parameters, Λ varies over the widest limits; Λ decreases as r, v, and ϱ increases, since this enhances the screening. The mean values of Λ and Γ are close to $\frac{1}{2}$.

The theory can be verified by observation. For example, if the observations give J* and v, and ϱ is known from independent observations, then by eliminating $\frac{dM}{dt}$ from (63) and (64), we obtain

$$4\pi J = \frac{\tau}{2} \Lambda S \varrho v^5. \tag{66}$$

* If m_{10} is the apparent magnitude reduced to the zenith and to a standard distance of 100 km, then

$$J = \text{const} - 0.4 m_{100}.$$

Öpik gives a constant of 9.72, if J is expressed in erg/sec and applies to the interval 4500 — 5700 Å.

Thus, it follows that the luminous intensity of a meteor should smoothly increase at first (mainly due to the increase in ϱ), and then should drop sharply (mainly due to the rapid decrease of v at the end of flight); this trend is actually observed (see Figure 145). The occasional deviations from the standard curve may be attributed to sudden changes of S due to fragmentation in flight (Figure 146).

The three equations, the braking, vaporization, and luminosity equations, can be solved for the three unknowns M, $\frac{dM}{dt}$, and ϱS, provided all the other quantities are known. If the density δ of the meteoric body is also assumed known, then from the mass M we can find the area S and the density ρ of the atmosphere. This accounts for the great geophysical significance of meteor observations.

In particular, M is usually determined from the integral form of the luminosity equation:

$$M = -8\pi \int_{t}^{t_d} \frac{J}{\tau v^2} dt, \tag{67}$$

where t_d is the moment of meteor disappearance, and M is its mass at the time t. The calculated masses have the following orders of magnitude. A meteor with $m_{100} = 0^m.0$ and an initial velocity v_0 of 40 km/sec directed downward along the vertical has an initial mass of $M_0 = 0.055$ g (Öpik's old scale) or $M_0 = 0.56$ g (Öpik's new scale).* The difference is due to the different values of τ_0 and p assumed.

The theory enables us to determine the conditions favorable for the falling of meteorites. If we divide (63) by (62) and integrate, we obtain

$$M = M_0 e^{-\frac{\sigma}{2}(v_0^2 - v^2)}, \tag{68}$$

where $\sigma = \frac{\Lambda}{2\Gamma Q}$. For ordinary meteors $\sigma = 10^{-12}$ sec^2/cm^2, and the residual mass remaining after the meteor has lost its space velocity is $M' = M_0 e^{-\frac{\sigma v_0^2}{2}}$; the latter is usually negligible, which justifies the application of the formula in the footnote on the preceding page.

For M' to be large, a high M_0 and a low v_0 are needed, and these are actually observed. Experience shows that meteorites fall only when the parent bolides (that is, fireballs, the brightest meteors in the sky) enter the atmosphere with velocities $v_0 < 11-12$ km/sec. More details are given on meteorites in Krinov's text (see reference list at end of chapter). The theory also makes it possible to determine the surface temperature T_0 of the meteoric body. In order to find this temperature, an additional heat-conduction equation is applied for high altitudes, before intensive vaporization sets in, and then this equation is subsequently replaced by equations (6) of §95. The result gives T_0, which increases to 2 or 3 thousand degrees in the lower part of the trajectory, in the zone of intensive vaporization.

§104. **The results of photographic observations of meteors.** Photographic observations of meteors received a very strong impetus in connection with the IGY—IQSY program. Such observations are made using special meteor patrols comprising high-speed-film cameras. Figure 140 shows one such

* In order to pass to other m_{100} and v_0 values, we may use the approximate logarithmic relation $\lg M_0 = \text{const} - 0.4\, m_{100} - 3 \lg v_0$, which follows from the formula in the footnote on the preceeding page and from (68), assuming constant τ_0 and equal dependences of J on v and t.

meteor-patrol unit from the Astrophysical Institute of the Tadzhik Academy of Sciences (cameras with Uran-9 objectives, 1/2.5, f = 250 mm, and a rotating shutter which interrupts the exposure to determine the meteor velocity).

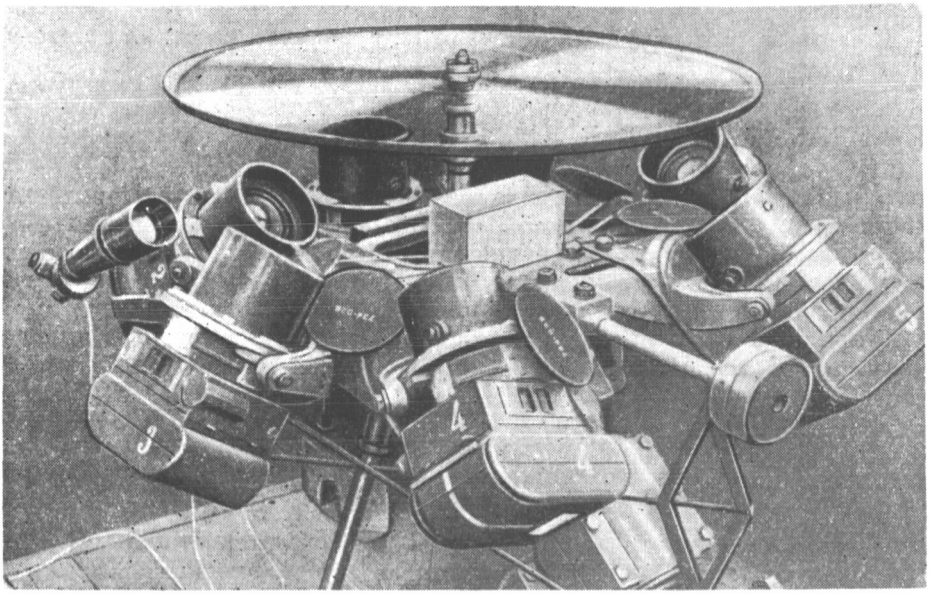


FIGURE 140

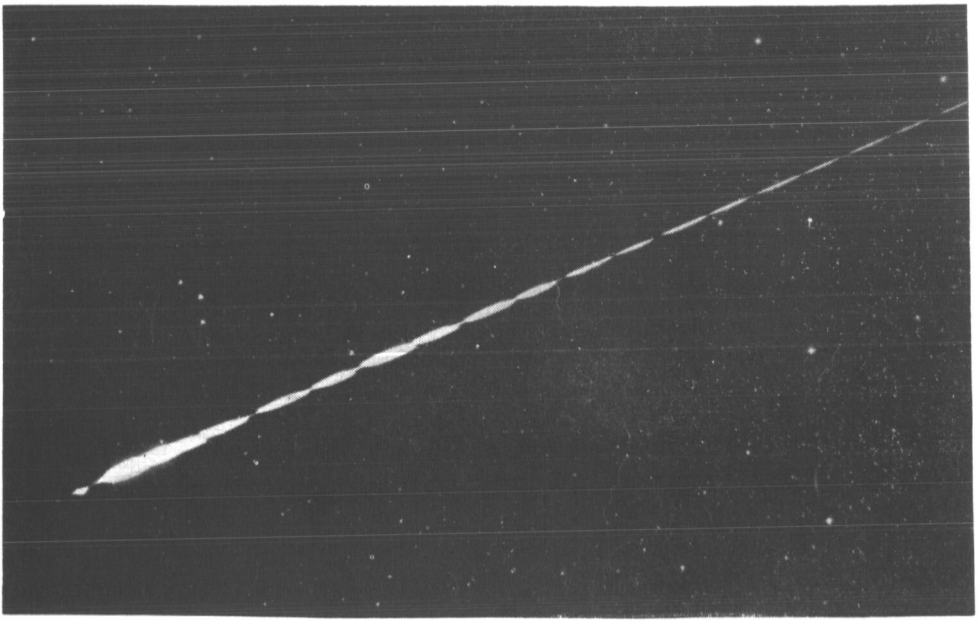


FIGURE 141

Observations from stations 30 or 40 km apart (Figures 141 and 142) make it possible to determine the altitudes, velocities, and accelerations of all the points on the visible meteor trail, and thence the characteristics of the atmosphere and of the meteor orbit. In recent years, very accurate catalogs of photographic orbits of meteors (containing over 500 orbits) have been compiled by Whipple and Jacchia in the USA, by Babadzhanov and Sosnova in the USSR, and by others. About 99% of the orbits in these catalogs are elliptical and of the quasiparabolic cometary type; asteroid-type orbits are almost not represented at all. About 1% are hyperbolic orbits, which apparently pertain to meteoric bodies of cometary origin whose motions have been especially perturbed. This seems to establish that the great majority of meteors are the products of fragmentation of cometary nuclei. Individual orbits have semimajor axes $a < 1$ A.U. and retrograde motion, but the origin of these is still not clear.

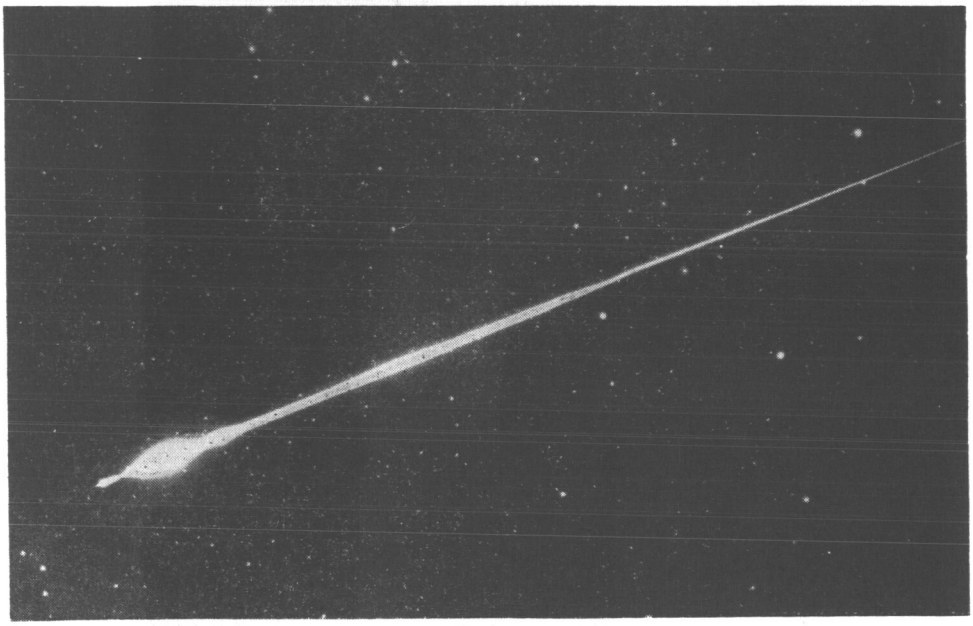


FIGURE 142

In addition to the famous large meteor streams, many weak streams have been discovered with radiants which are scattered over the celestial sphere, like the radiants of sporadic meteors. This confirms the opinion that sporadic meteors are in fact the last stage in the dissipation of weak streams. The appearance and disappearance altitudes H_1 and H_2 and the average geocentric velocities v_g of meteors have been determined accurately (see Figure 143); also, the dependences of H_1 (dots) and H_2 (crosses) on the absolute magnitude m_{100} have been determined (Figure 144).

Ivanikov employed a meteor-simulation laboratory unit to study in detail the errors arising during meteor photometry.

From the photographic curves of 392 meteors, Jacchia determined the altitudes h at which individual meteors reach a photographic magnitude of

$+2^m,5$. He found that more massive, brighter meteors reach this magnitude at higher altitudes; in general, h increases by 1.5 km for each stellar magnitude of the meteor. Jacchia also studied in detail the influence which fragmentation effects have on the nature of the photographic curve of a meteor (Figures 145 and 146).

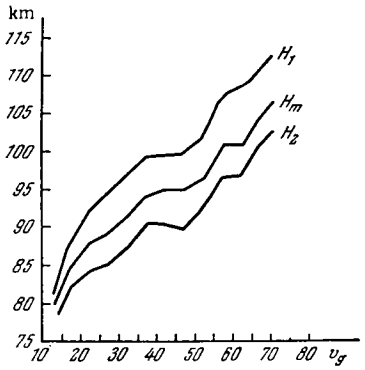


FIGURE 143

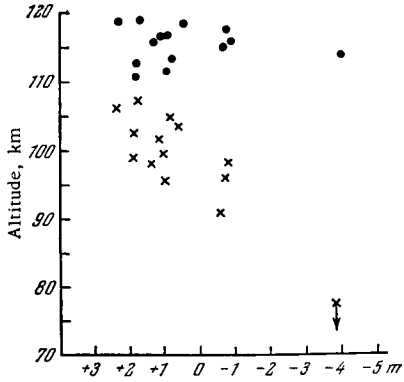


FIGURE 144

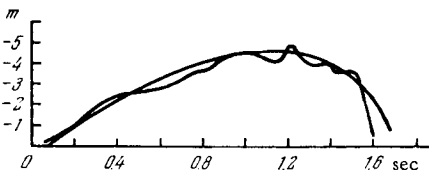


FIGURE 145

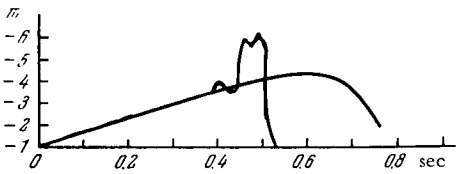


FIGURE 146

Direct data have been obtained on the fragmentation of meteoric bodies. 25% of the weak meteors photographed with a Superschmidt camera (1/0.85, f = 20 cm) have blurred shutter breaks due to swarms of particles separated from the meteoric body. A unique photograph was obtained of a bolide followed by a meteorite (the Przybram chondrite, Czechoslovakia), which broke up into 17 bodies at an altitude of 44 to 23 km.

Using the 48" Schmidt camera at Mount Palomar Observatory, Cook, Hawkins, and Whipple determined the trail widths for meteors between $+1^m$ and $+4^m$; these were found to be around 1.3 ± 0.5 m (and 3 m for Geminids). One of the 0^m meteors was much wider, separating toward the end point into three parts which diverged by 20m.

Many determinations have been made of the atmospheric parameters. It was found, however, that the atmospheric densities obtained by meteoric methods differ systematically from the values obtained by rocket techniques, apparently due to the excessive values assumed for the densities of meteoric bodies. By reversing the calculation, that is, by assuming a value of ϱ from rocket observations and attempting to calculate the density δ of meteoric

bodies, Whipple found that as a rule $\delta \ll 1$, and that in some cases it dropped to 0.05 g/cm^3. These estimates are consistent with the "ice" model of the cometary nucleus, but they do not agree very well with the results obtained for dust tails (§101); δ apparently differs for different meteor streams, and some improvement in the determination of this quantity can be expected in view of the current reconsideration of the basic constants of the physical theory. The most probable mean value for δ is assumed to be ~ 1 g/cm^3.

§105. **Spectra of meteors.** The luminosities of the meteor itself, of the meteor tail (as observed during shutter breaks), and of the meteor trail, mainly show up as a gaseous cloud surrounding the meteoric body. So far, over three hundred meteoric spectra have been obtained, plus individual spectra of tails and trails; these were obtained mostly in Canada (Millman), Czechoslovakia (Rajhl, Cepleha), and the USSR (Lyubarskii, Ivanikov).

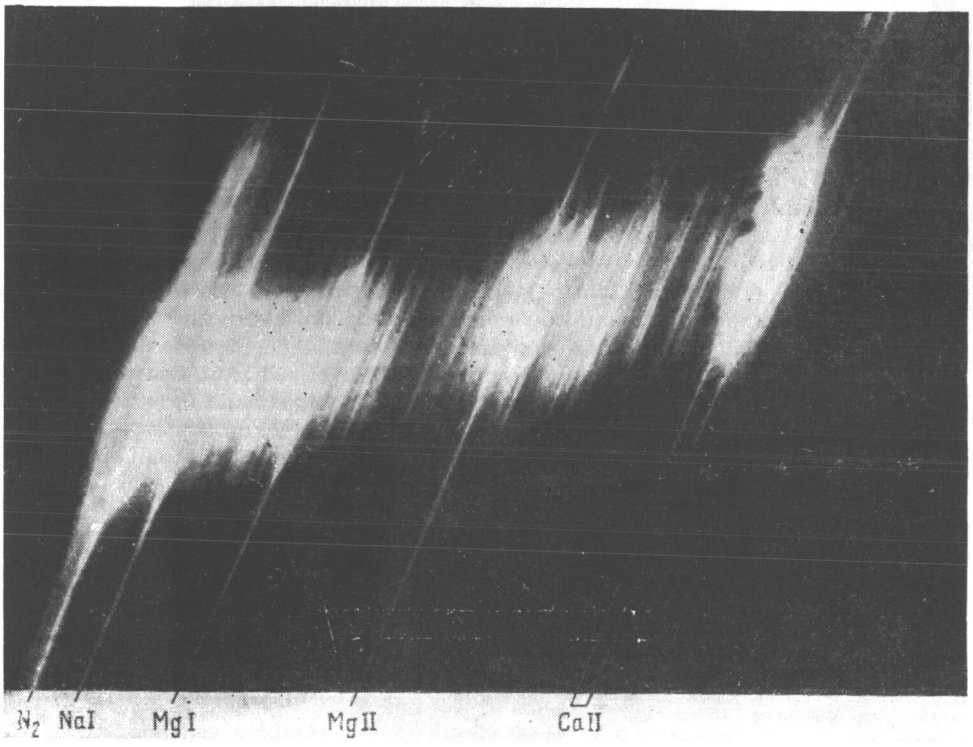


FIGURE 147

The spectra of meteors contain atomic emission lines, mostly those of the multiplets CaI, CaII, FeI, FeII, NaI, MgI, MgII, SiI, SiII, MnI, CrI, NiI, and also HI. Onto these lines is superposed a continuum, partly due to the molecular emission bands of atmospheric nitrogen: due to the first positive system ($B^3\Pi_g \rightarrow A^3\Sigma_u^+$ transition) in the red region and due to the second positive system ($C^3\Pi_u \rightarrow B^3\Pi_g$ transition) in the blue-violet. Figure 147 gives the spectrum of a Perseid meteor taken on 12 August 1958. The forbidden green line of OI (5577 Å) was also observed. The spectra of meteor tails and trails contain only the lines of the lowest excitation states of Fe, Ca, CaII, Na, and Mg.

The best spectrograms, with dispersions of 15 — 25 Å/mm, include over 250 lines (Rajhl and Cepleha at Ondjeiov Observatory, Czechoslovakia).

TABLE 13

Stream	v_g, km/sec	Nature of spectrum
Giacobinids	23	Low excitation, no lines of ionized elements
Geminids	36	Very weak CaII emissions
Perseids	60	Strong emissions of CaII, MgII, SiI
Orionids	68	
Leonids	71	The same; MgII emissions particularly distinct, imparting to the meteors the characteristic greenish hue.

The nature of the spectrum is mainly determined by the geocentric velocity v_g of the meteor and, to a lesser degree, by the chemical composition of the meteoric body (see Table 13). Millman suggests the following classification of meteor spectra: type *Y*, in which the H and K CaII emission lines predominate; type *Z*, in which most of the lines are iron lines, with CaII absent; and type *X*, an intermediate type. The effective temperature T_{eff} of the meteor is determined as the temperature in a King furnace required to produce the same relative intensity of the Fe multiplets as in the meteoric spectrum. It should be remembered, however, that in meteors thermodynamic equilibrium is completely absent. T_{eff} depends essentially on v_g and varies on the average from 1700°K for v_g = 12 km/sec to 3000°K for v_g = 71 km/sec; for a given value of v_g, it is higher the brighter the meteor and the lower the meteor altitude. Color indexes were determined by Millman by integrating the luminous intensities in the photographic ($\lambda\lambda$ 3700 — 5000) and visible ($\lambda\lambda$ 5000 — 6700) spectral regions and by comparing the results with the values found for AO stars (the comparison star used was Vega). It was established that for type-*Z* meteors the color index is close to zero, while type-*Y* meteors have on the average negative color indexes which vary greatly with the brightness: they are up to $+1^m$ for faint meteors and up to $-1^m.5$ for bright meteors with strong H and K lines of CaII.

Many meteors have variable spectra. For these, the excitation is generally low at the beginning of the path and increases toward the end point, where the brightness is a maximum. The temperature T_{eff} and the color change correspondingly (in the Perseids, for example, from yellow to blue). All the spectra of one meteor stream are close to one another. No systematic differences are observed between the spectra of shower and sporadic meteors. The various observed features of meteoric spectra conform on the whole to Opik's theoretical interpretation, according to which the meteor luminosity is due to impact-excited (direct collisions of air molecules with the atoms of the gaseous coma of the meteor) and temperature-excited (molecular collisions in the coma) emissions. Impact excitation predominates in fast meteors. Temperature excitation predominates in slow meteors in the lower parts of their paths, where the density of the gas coma is quite high, particularly in the shock-wave zone. The coma, according

to Opik, is underexcited; the brightest lines are those originating in the lower atomic levels. Impact excitation produces the lines of various ionized states, while temperature excitation gives only the lines of singly ionized states of low-ionization atoms. Table 14 lists the typical spectral lines of a Perseid meteor, according to Millman.

TABLE 14

λ	First-order equivalent width, A	Relative intensity	Atom or ion	Excitation potential, eV	Relative abundance of atoms in upper state
3934, 3967	26	150	Ca II	3.1	180
4061	3	11	Fe I	4.6	10
4092	2	9	Si I	4.9	
4140	1	5	Fe I	4.5	
4196	1	5	Fe I	4.4	
4226	2	5	Ca I	2.9	3
4268	2	4	Fe I	4.4	
4321	4	5	Fe I	4.4	
4388	4	6	Fe I	2.8	
4425	5	6	Fe I	2.8	
4483	11	11	Mg II	11.6	10
4594	2	2	Fe II	5.5	
4714	1	2	Mg I	6.9	20
4920	2	2	Fe I	3.3	
5037	2	3	Fe I	3.3	
5176	13	16	Mg I	5.1	30
5261, 5319, 5424	9	8	Fe I	3.2	600
5588	1	1	Ca I	4.6	2
5687	2	1	Si I	7.1	
5893	34	18	Na I	2.1	80
6153	3	1	Fe II Na I	5.9 4.1	
6249	3	2	Fe II	5.9	
6359	23	9	Fe II	5.8	40
6565	1	1	H I	12.0	

§106. Radar observations of meteors. A meteor trail is always ionized. The main causes of this are: impact ionization, photoionization, detachment of ions due to cathode sputtering in the upper atmospheric layers, and thermal ionization in the lower layers. Radar techniques can therefore be applied in meteor research.

The mechanical characteristics (radiants, velocities, accelerations) and physical characteristics (electron concentrations in meteor trails, rates of diffusion) of meteors can be determined by means of radar. The methods of radiant determination are based on the principle of mirror reflection: the radar picks up a reflected signal only when the radio beam is perpendicular to the trail.

Meteor velocities are determined in three ways:

1. By reflection from the head part of the trail (from the ion cloud surrounding the meteoric body). Unlike the main reflection, this is observed for different inclinations of the radio beam to the trail, but only for the brightest meteors. The ion cloud apparently forms due to ionization of the molecular oxygen in the air by the ultraviolet radiation of the meteor.

2. By the diffusion of radio waves from the trail. This phenomenon is analogous to the Fresnel diffraction at a knife edge which is observed in optics. The knife edge, in this case the meteor trail, moves relative to the observer, who alternately receives diffractional maxima and minima.

3. By the method of undamped waves. A diffraction pattern is observed when an undamped wave sent by the transmitter interacts with a wave scattered from the meteor trail.

Radiophysical methods of research are based on the so-called fundamental radar equation. This equation defines the total energy supplied to the first stage of the radar receiver per unit time as

$$E=\frac{P}{4\pi R^2}G\frac{\sigma}{4\pi R^2}\frac{G\lambda^2}{4\pi}=\frac{PG^2\lambda^2\sigma}{64\pi^3R^4} \tag{69}$$

Here P is the transmitter power, R is the target distance, G is the directional-effect coefficient of the antenna, σ is the equivalent reflecting surface of the target lying in the plane of the main lobe of the antenna, λ is the radar wavelength, and $\frac{G\lambda^2}{4\pi}$ is the effective absorbing area of the antenna.

If the target is a single electron, then $\sigma = 4\pi r_0^2$, where $r_0=\frac{e^2}{mc^2}$ is the classical radius of the electron. An analogous expression applies for positive ions, but since the mass of ions [present in the meteor trail?] is three orders of magnitude less than the mass of electrons, the ions need not be taken into account.

A meteor leaves a trail in the form of a thin ion column. In order to determine the corresponding E, we must add up the reflections from all the electrons of the trail, taking into consideration phase differences due to the different distances of the various trail elements from the receiver. This was done by Lovell and Clegg for a linear electron density in the meteor trail (that is, the number of electrons per cm of trail length) of $\alpha < 10^{12}$ particles/cm (unsaturated trail). These authors obtained the following expression for the energy:

$$E=\frac{\alpha^2PG^2\lambda^3r_0^2}{12\pi^2R^3}\text{ watts.}^* \tag{70}$$

It was assumed here that the electric vector was parallel to the trail, and that the trail width was $\ll \lambda$.

For $\alpha > 10^{12}$ particles/cm (saturated trail), the reflection is like that from a metallic cylinder, so that, according to Gloos and Greenhow,

$$E=\frac{\alpha^{1/2}PG^2\lambda^3r_0^{1/2}}{20\pi^3R^3}\text{ watts.} \tag{71}$$

Once the trail forms, it starts to diffuse immediately, and the initial amplitude A_0 of the signal begins to decrease exponentially:

$$A=A_0\exp\left(-\frac{16\pi^2Dt}{\lambda^2}\right), \tag{72}$$

where D is the diffusion coefficient for $\alpha < 10^{12}$ particles/cm.

In the limiting case when D is very large, while the velocity of the meteor (and therefore the rate of production of the ionized trail) is low, the initial amplitude A_0 cannot be established. When $\alpha > 10^{12}$ particles/cm, the signal duration is

$$\Delta T=\frac{r_0\alpha\lambda^2}{4\pi^2D}. \tag{73}$$

Formulas (71) and (73) enable us to determine, from the measured E, the linear concentrations α and the diffusion coefficient D for bright meteors,

* Unit of length, meter.

while formulas (70) and (72) apply to faint meteors. For a typical faint meteor ($+5^m$) falling vertically with a velocity of 40 km/sec, we find that $E \approx 10^{-11}$ watt and $\alpha \approx 10^{10}$ electrons/cm; for a bright meteor (-5^m), we have $\alpha \approx 5 \cdot 10^{14}$, while D varies with the atmospheric density, increasing exponentially with the height of the meteor.

In addition to the three fundamental equations of the physical theory of meteors, (62), (63), and (64), meteor radio astronomy also makes use of the ionization equation

$$\alpha = \frac{dM}{dt} \frac{1}{v} \frac{\beta}{\mu}, \tag{74}$$

where μ is the mass and β is the ionization probability of a vaporized meteoric atom. Given α, we can find $\frac{dM}{dt}$ from (63) for known J, and then determine β. The probability β can also be computed theoretically on the basis of the quantum-mechanical theory of the collision of slow particles ($v \ll c$). In the range from 20 to 70 km/sec, apparently $\beta \sim v^3$; for $v = 40$ km/sec, $\beta = 0.14$; for $v \leqslant 20$ km/sec, β starts to decrease rapidly; and for $v = 15$ km/sec, $\beta \approx 0.0006$.

If β is known, we can find the density of the atmosphere independently: α is available from observations, $\frac{dM}{dt}$ is calculated from (74), and the density ϱ can then be found from (63).

Radio-astronomical studies of meteors have developed rapidly in the last decade. The main centers of research are: the USSR (Khar'kov, Tomsk, Kazan'), Canada (Springhill, Saleatun, Minook, Baker Lake, Resolute Bay), Czechoslovakia (Ondjeiov), England (Jodrell Bank), The USA (Havana, Illinois), Australia (Adelaide), and Japan (Tokyo).

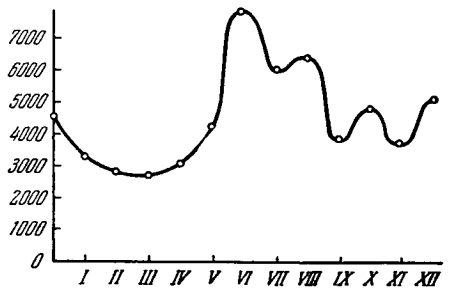


FIGURE 148

Radar observations are distinguished by their high capacity. For instance, the Havana bank, operating on a 7.2 m wavelength with a peak power of 7 mw, records up to 1000 meteors per hour, down to meteors of 12th stellar magnitude. Thus, many millions of observations have been accumulated so far. This vast amount of data makes it possible to obtain reliable meteor statistics, to determine the seasonal and daily variations of meteor activity. Figure 148 shows the seasonal variation; the ordinate indicates the monthly average of the daily rate. Figure 149 shows the daily variation, with a maximum near apex culmination (both curves were plotted according to the data of the Khar'kov Polytechnical Institute). The statistical

data available make it possible to calculate the average orbital elements of meteor streams with the same accuracy as photographic techniques (although the individual orbits remain much less accurate), and to investigate with considerable success the structure of the meteor streams. For example, the stream cross sections of the Perseids have an almost uniform structure, which is evidence of old age; the Quadrantids, conversely, have a large number of small meteoric bodies in the middle of the stream, which have not yet dispersed under the action of corpuscular bombardment and radiation pressure (the Poynting-Robertson effect). An exceptionally concentrated jet of meteoric particles causing heavy, but short-lived showers was observed in the Lyrids, which in this respect was similar to the very narrow stream of Monocerotids, which reaches a maximum on 21 November with a visual hourly rate of up to 3000.

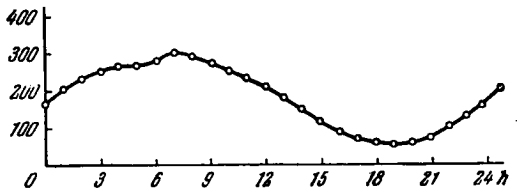


FIGURE 149

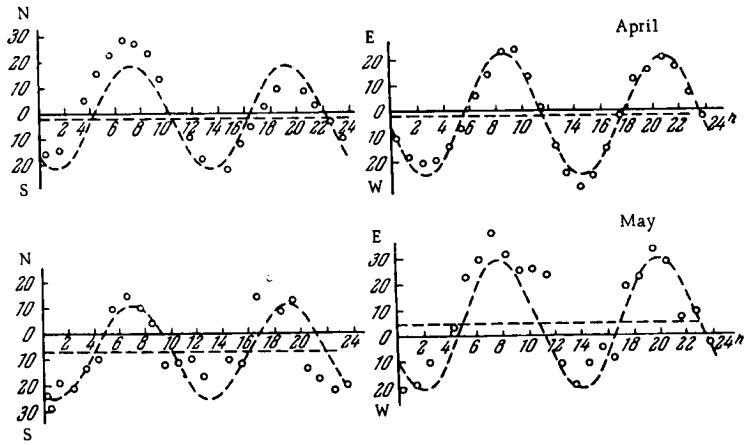


FIGURE 150

Statistical data on the drift of ionized trails give a detailed picture of the prevailing winds at the altitudes of meteor appearance. The basic 12-hour period and the superimposed 24-hour period with half the amplitude have been established. Figure 150 shows the diurnal variations of NS and EW wind components (m/sec) for the months of April and May 1960. The abscissas give the local solar time (Khar'kov); the circles mark the mean hourly values, and the dashed lines show the 12-hour wind harmonics. Figure 151 shows polar diagrams of the 12-hour harmonic wind component:

a is the diagram for the regions E and E_{spor} for March, April, and May; *b* and *c* are diagrams plotted from the harmonics in Figure 150 for April and May, respectively (1960). Figure 152 shows the dependence of the initial trail width r_0 and the diffusion coefficient D on the elevation above sea level.

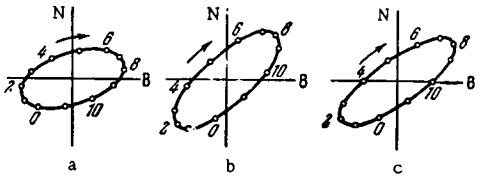


FIGURE 151

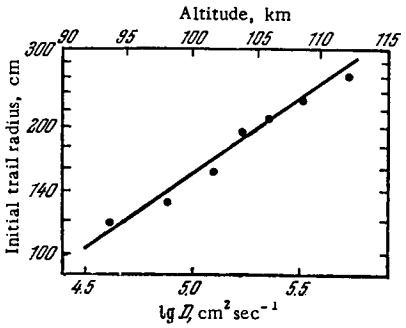


FIGURE 152

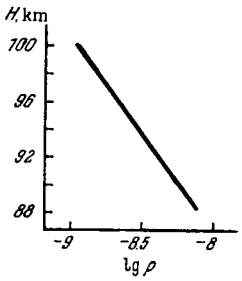


FIGURE 153

The altitude as a function of meteor velocity was found to be the same as for photographic meteors. The variation of atmospheric density with altitude is shown in Figure 153.

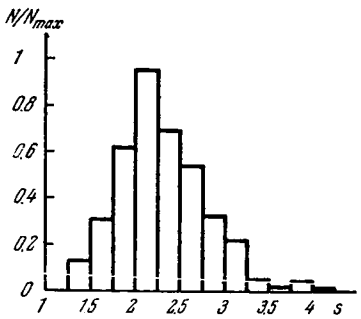


FIGURE 154

The mass distributions of meteors have been studied in detail. The mean value of the exponent s in the distribution law was found to be close to 2, just as for visual observations (Figure 154).

§107. The latest research techniques. The implementation of the IGY-IQSY program resulted in the introduction, mainly in the USSR and the USA, of large-scale rocket and satellite studies of meteors. Some very diverse measuring procedures were tried: pit counts on soft metal and on polished steel, piezomicrophones, photoelectric recordings of impact flashes in plexiglas or in an aluminized lucite disk, the resistance increase in a coated metallic layer, the radioactivity decrease in a standard, etc. As a result, particles with minimum energies of $\sim 10^4$ eV (or minimum masses of 10^{-9}g for velocities of 40 km/sec) could be detected. The results of the different measuring techniques can hardly be compared at this stage; however, they have already led to some important discoveries. For instance, a dust belt was discovered, surrounding the earth up to altitudes of $\sim 10^5$ km and gradually merging with the interplanetary dust background (the zodiacal cloud, see § 108 and §109). Figure 155 gives the number of impacts N of meteoric bodies (mass, 10^{-9}g) per m^2 per sec as a function of the distance h (km) from the earth. At the appearance heights for meteors an atmospheric aerosol layer was discovered;; this constantly present layer extends over altitudes between approximately 80 and 100 km, with maximum densities at 85 km in the middle latitudes and at 92 km in the northern latitudes. At the same altitudes, mass spectroscopes detected a thin layer of metal atoms, most likely meteoric decay products.

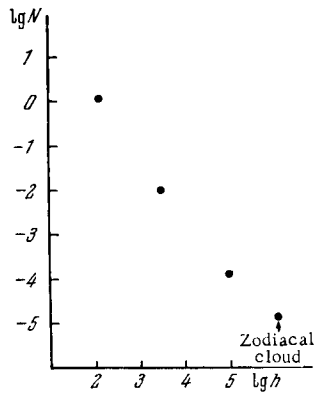


FIGURE 155

In June 1961 a rocket fitted out with Mylar nitrocellulose, lucite, and other plastic strips collected a considerable number of micrometeorites, which were later studied in the laboratory. Most of the trapped particles were of submicron size, and they fell into three groups: 1) high-density spheres; 2) medium-density irregular particles; 3) highly irregular particles of average density ("flaky" particles). Electron-diffraction methods, electron excitation of characteristic X rays, and neutron activation were used to determine the chemical composition of the micrometeorites. Different particles were found to have different chemical compositions.

For example, some particles had high contents of Al, Si, and Fe, and hardly any Ca and Mg, while others were rich in Ti and Ca, poor in Fe and Si, and contained almost no Mg. Iron-nickel sulfide $(FeNi)_3S_4$ was identified definitely.

Recently, a new photoelectric method of meteor recording was developed, and in the future it should replace low-accuracy photographic photometry. As an addition to the laboratory experiments, some first successful trials have been made of simulating meteors by means of ejecting bodies into the free atmosphere from rockets.

§108. General data on the zodiacal light. In the middle latitudes the zodiacal light is observed after twilight in the western sky from September to May and before dawn in the eastern sky from the end of July to the middle of March. It is cone-shaped, with its base extending along the horizon 15 to 20° to either side of the ecliptic and with a height of 50 to 60°.

As the sun sinks below the horizon, the zodiacal light wanes. In the middle latitudes it is generally visible up to elongations ε of about 90° from the sun. At low latitudes, where the visibility is more favorable, the zodiacal light can be observed up to elongations of $\varepsilon = 180°$. The evening and the morning cones merge in this case, forming the zodiacal band. The intensity of the morning zodiacal light is a maximum in January and February, and that of the evening zodiacal light is a maximum in April and May; during these periods the zodiacal light is appreciably brighter than the light of the Milky Way.

The apparent axis of the cone departs somewhat from the ecliptic. In the northern hemisphere it is a few degrees to the north, and in the southern hemisphere it is a few degrees to the south; this phenomenon is mainly due to atmospheric extinction. According to recent measurements made on high mountains, the true plane of symmetry of the zodiacal light lies in the invariable plane of the planetary system (determined by the direction of the total angular momentum of the planetary system), and it meets the ecliptic at a longitude of 115°, making an angle of $\sim 1°.5$ to it.

Photometric analyses of the zodiacal light were undertaken by various authors. The most reliable observations are those made at high altitudes, where the atmospheric interference is less. The results of recent high-altitude observations led to the graph in Figure 156, which clearly depicts the gradual merging of the solar corona with the zodiacal light. The left-hand ordinate gives the surface brightness H in units of solar-disk brightness $B_\odot$ multiplied by 10^{-12}. The right-hand ordinate is the degree of polarization P of the light: 1) of the inner corona according to van de Hulst's modified model; 2) of the outer corona according to Blackwell's observations during the eclipse of 30 June 1954 from a height of 10 km; and 3) of the zodiacal light according to the 1958 observations of Blackwell and Ingham in Chacultaya (Bolivian Andes) at a latitude of 16°S and an altitude of 5.1 km. The reddening of the zodiacal light relative to the sun according to the Chacultaya observations is illustrated in Figure 157, which gives H in units of $10^{-13}B_\odot$ for $\varepsilon = 35°$ in three spectral regions.

H fluctuates irregularly from night to night and even from hour to hour. The sharpest fluctuations are generally associated with chromospheric flares and magnetic storms (see Figure 158, which correlates the increments of H with the magnetic index K, according to the Chacultaya observations).

The spectrum of the zodiacal light is a solar Fraunhofer spectrum with telluric emissions superimposed onto it (Figure 159). The so-called counterglow is apparently an independent phenomenon. This shows up as a weak increase (by 13%, on the average) of the night-sky brightness at a point diametrically opposite to the sun. The counterglow has the shape of a roughly elliptical spot with its major axis lying almost in the plane of the ecliptic. The axes reach to about 10° and 6°, and under favorable conditions to 40° and 10°. The center of the spot is displaced somewhat toward the west, relative to the counter-solar point. The counterglow is sometimes deformed, and extends westward in the form of a fan-shaped diverging tail or cone, not unlike the zodiacal light. Such a distorted counterglow is called the false zodiacal light.

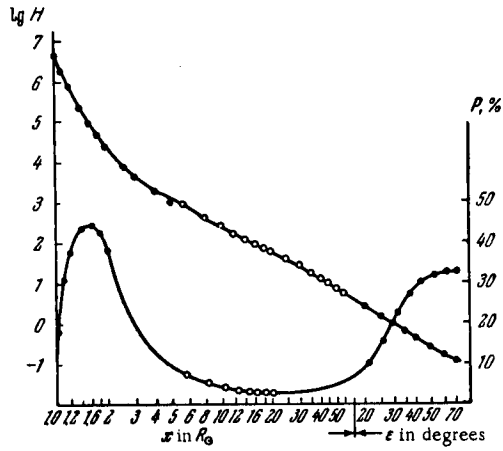


FIGURE 156

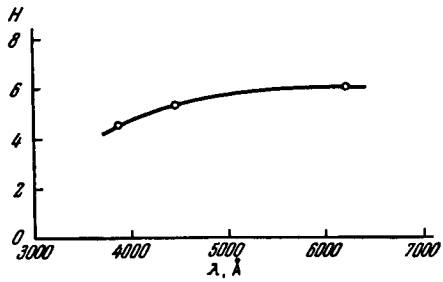


FIGURE 157

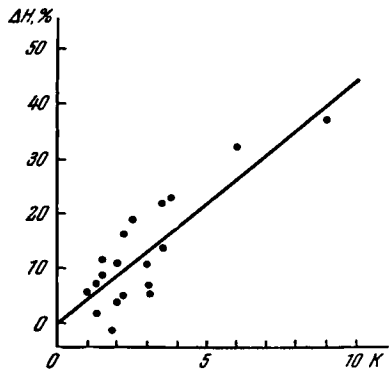


FIGURE 158

The counterglow parallax was first discovered by Astapovich. According to the measurements of Rozhkovskii, the horizontal parallax of the counterglow is 3°.5.

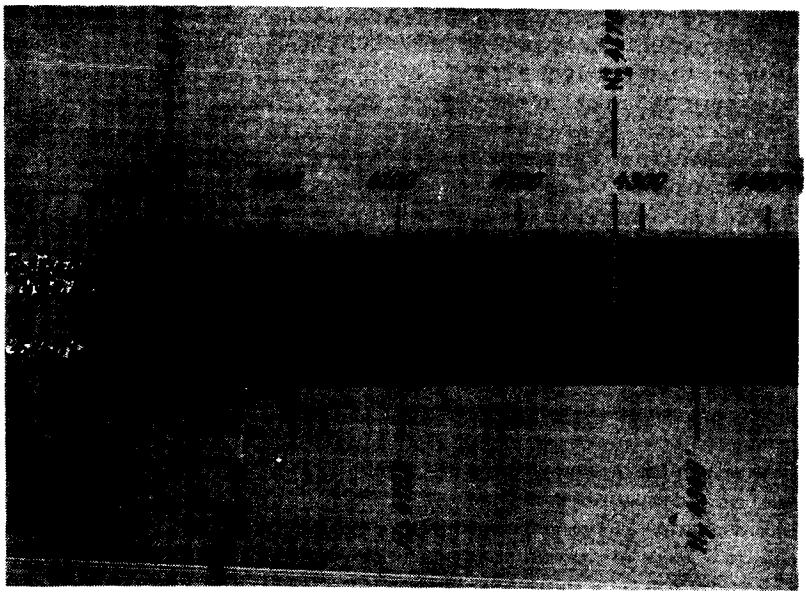


FIGURE 159

§109. The nature of the zodiacal light. At present, the great majority of observers regard the zodiacal light as a cloud of scattering particles around the sun. This cloud is lens-shaped and lies in the primary (invariable) plane of the planetary system. The object of theory is to determine, from the apparent glow of the zodiacal light, the nature of the scattering particles, their size-distribution function, and their volume density.

Since all these problems are difficult to solve simultaneously, additional assumptions have to be introduced. Correspondingly, different models of the zodiacal light have been suggested. The common premises of all these models are the following:

1) The scattering centers are mainly meteoric particles of density δ with radii $r > r_0 = 5.7 \cdot 10^{-5} \delta^{-1}$ cm (smaller particles would have been removed by the radiation pressure of the sun, by analogy with the particles in cometary tails).

2. The zodiacal light is an extension of the Fraunhofer corona of the sun.

3. The zodiacal light is continually replenished as particles fall into the sun and as new meteoric matter is added from the planetary system by means of radiation braking (the Poynting-Robertson effect). The radiation braking is produced as follows: due to the aberration of light, the radiation-pressure force departs somewhat from the radius vector, and thus has a tangential component which is proportional to the linear orbital velocity of the particle. Calculations show that particles with initially circular orbits of radius a fall into the sun after $t = 7 \cdot 10^6 r\delta a^2$ years (r is in centimeters, and δ is in g/cm^3); particles with elliptical orbits will fall more rapidly, for example, the corresponding value for the Geminid orbit ($a = 1.4$, $e = 0{,}9$) is $t = 1.4 \cdot 10^6 r\delta$ years.

The characteristic features of the various models will now be outlined.

a) Fesenkov extrapolated the distribution function for meteors (61) to r_0 inclusive, setting 1 cm $\geqslant r \geqslant r_0$; the albedo was arbitrarily taken as 1, and $\delta = 3\,g/cm^3$. He assumed that the volume density of the dust particles $D \sim R^{-1}$ (where R is the distance from the sun), which corresponds to a stationarity condition in the presence of radiation stopping. Fesenkov calculated that the observed brightness of the zodiacal light can be attained if the density near the earth is $D_0 = 7.6 \cdot 10^{-23}\,g/cm^3$.

b) Allen made calculations for $r = 10^{-2}$, 10^{-3}, and 10^{-4} cm and for $D = \text{const}$, R^{-1}. R^{-2}; he found that the best agreement with observations is obtained for $r = 10^{-3}$ cm and $D = \text{const} = 2 \cdot 10^{-23}\,g/cm^3$.

c) van de Hulst assumed $D = \text{const}$ for $\varepsilon < 30°$. Proceeding from the theory of light diffraction by small dust particles and using the empirical relation $H \sim \varepsilon^{2.3}$, he obtained $\varphi(r) = r^{-2.6}$. In this case, $D = 5 \cdot 10^{-21}\,g/cm^3$ fits the observations best.

d) Behr and Siedentopf suggest that the polarization of the zodiacal light may be due to free electrons. By setting $r = 10^{-3}$ cm and attributing the polarization to electrons only, they found $D = \text{const} = 10^{-23}\,g/cm^3$; the electron concentration $n_\odot$ at a distance of 1 A.U. from the sun was found to be 600 electrons/cm^3. This value, however, is inconsistent with present-day spectroscopic data and must therefore be rejected. For example, the Chacultaya observations gave an upper limit of $n_\odot = 120$ electrons/cm^3.

Other authors (Elsasser, Minnaert, Siedentopf), assumed various values for D and φ and obtained values for D_0 ranging from $2 \cdot 10^{-23}$ to $6 \cdot 10^{-22}\,g/cm^3$.

Finally, Ingham, assuming an electron density of

$$n_\odot(r) = 6.75 \cdot 10^6 R^{-3} \quad (R \geqslant 6),$$

where R is in solar radii, found that the Chacultaya observations can be represented by various functions $\varphi(r)$ in which the modal (most frequently occurring) value of r lies near the lower limit. Depending on the actual parameters of $\varphi(r)$, D_0 ranges between 0.33 and $5.3 \cdot 10^{-24}\,g/cm^3$, which corresponds to volume densities of 600 to 17 particles per km^3. The albedo of the particles varies between 0.16 and 0.56.

Attempts have been made to explain the counterglow as an accumulation of meteoric particles near the Lagrangian point of libration L_2 of the restricted three-body problem (sun, earth, and meteoric body), which lies on the continuation of the earth's radius vector (Gulden, Moulton). However, no satisfactory agreement with observations could be obtained. A better fit is achieved using the theory of an atmospheric origin for counterglow, which was advanced by Zakharov and developed by Astapovich and Fesenkov. According to this approach, the earth's atmosphere forms a rarefied, slightly divergent tail almost parallel to the radius vector, similar to a cometary tail. The effects associated with the glow of this tail as viewed against the night sky provide qualitative explanations of both counterglow and the false zodiacal light.

More reliable information on the nature of the zodiacal light will be provided by measurements from space rockets located outside of the earth's dust belt.

REFERENCES FOR CHAPTER XIV

1. AMBARTSUMYAN, V.A. Teoreticheskaya astrofizika (Theoretical Astrophysics). — Gostekhizdat. 1939.
2. ASTAPOVICH, I.S. Meteornye yavleniya v atmosfere Zemli (Meteor Phenomena in the Earth's Atmosphere). — Fizmatgiz. 1958.
3. BABADZHANOV, P.B. and E.N. KRAMER. Metodika i nekotorye rezul'taty fotograficheskikh issledovanii meteorov (Methods and Some Results of Photographic Investigation of Meteors). — In: "Rezul'taty issledovanni po programme MGG," Izdatel'stvo Akademii nauk SSSR. 1963.
4. BLACKWELL, D.E. and M.F.INGHAM. Observations of the Zodiacal Light from a Very High Altitude Station. — M.N. R.A.S. 122:2. 1961.
5. BREDIKHIN, F.A. O khvostakh komet (On Cometary Tails). — Gosudarstvennoe izdatel'stvo tekhniko-teoreticheskoe literatury. 1934.
6. DOBROVOL'SKII, O.V. Nestatsionarnye protsessy v kometakh i solnechnaya aktivnost' (Nonstationary Processes in Comets and Solar Activity). — Trudy Instituta Astrofiziki, Vol.8. Dushanbe. 1961.
7. FIALKO, E.I. Radiolokatsionnye metody nablyudeniya meteorov (Radar Methods for Meteor Observation). — Sovetskoe radio, Moskva. 1961.
8. HULST, H.C. VAN DE. Scattering of Light by Small Particles. — Izdatel'stvo inostrannoi literatury [Russian translation]. 1961.
9. IVANIKOV, V.I. O metodakh fotograficheskoi fotometrii meteorov (Photographic Photometry of Meteors). — Byulleten' Instituta astrofiziki, No.21. Dushanbe. 1957.
10. JÄGERMAN, R. Bredichins mechanische Untersuchungen an Cometenformen. — St. Petersburg. 1903.
11. KASHCHEEV, B.L. and V.N. LEBEDINETS. Radiolokatsionnye issledovaniya meteornykh yavlenii (Radar Investigations of Meteor Phenomena). — Rezul'taty issledovanii po programme MGG, Ionosfera i meteory, No.7. 1961.
12. KATASEV, L.A. Fotograficheskie metody meteornoi astronomii (Photographic Methods of Meteor Astronomy). — Gostekhizdat. 1947. (Also see BABADZHANOV, P.B. and A.K.SOSNOVA, — Byulleten' Instituta astrofiziki, No. 29. Dushanbe. 1960.)
13. KRINOV, E.L. Osnovy meteoritiki (Principles of Meteoritics). — Goztekhizdat. 1955.
14. LEVIN, B.Yu. Fizicheskaya teoriya meteorov i meteornoe veshchestvo v solnechnoi sisteme (Physical Theory of Meteors and Meteoric Matter in the Solar System). — Izdatel'stvo Akademii nauk SSSR, 1956.
15. LOVELL, A.C.B. Meteor Astronomy. — Fizmatgiz. 1958 [Russian Translation].
16. MAROCHNIK, L.S. Magnitogidrodinamicheskie yavleniya v kometakh, ikh svyaz' s geoaktivnymi potokami (Magnetohydrodynamic Phenomena in Comets and their Relation to Geoactive Streams). — Uspekhi fizicheskikh nauk. 1964.
17. Meteory (Meteors). — Collection of papers [Russian Translation]. Izdatel'stvo inostrannoi literatury. 1959.
18. MILLMAN, P. An Analysis of Meteor Spectra. — Harv. obs. Ann. 82: 6 (7).
19. ÖPIK, E. Physics of Meteor Flight in the Atmosphere. — New York. 1958.
20. ORLOV, S.V. Komety (Comets). - Ob"edinenie nauchno-tekhnicheskikh izdatel'stvo. 1935.
21. POLOSKOV, S.M. Fizika komet (Physics of Comets). — Dissertation.
22. RICHTER, N. Statistik und Physik der Kometen. — Leipzig. 1954.
23. SWINGS, P. and L. HASER. Atlas of Representative Cometary Spectra. — 1956.
24. VSEKHSYVATSKII, S.K. Fizicheskie kharakteristiki komet (Physical Characteristics of Comets). — Fizmatgiz. 1958. *

* [English Translation: Israel Program for Scientific Translations, Cat.No. 808. 1965.]

PRECEDING PAGE BLANK NOT FILMED.

INDEX

ABBOT 6—8, 10, 12, 14
Abel equation 81
ABETTI 58, 59
ABHYANKAR 51, 67
ADAMS 21, 303
AIRY 111
Albedo 241, 243, 247, 249, 285—288, 300, 315, 321, 326, 331
ALDRICH 8
ALEKSANDROV 263
ALFVÉN 66, 106, 110, 146, 363—365
ALLEN 6, 22, 26, 31, 32, 35, 78, 101, 109, 137, 147, 152, 390
ALTER 257
AMBARTSUMYAN 1, 36, 67, 264, 293, 317, 327, 363, 391
AMENITSKII 263
ÅNGSTRÖM 8, 20
Antennas 191
ARAGO 111
ASHBROOK 308
ASTAPOVICH 370, 388, 390, 391
Asteroids 2, 279, 325, 326, 328—335, 336
Astronomical unit 3
ATHAY 76, 89, 93, 94, 95, 98, 109
AUWERS 4

BAADE 323
BABADZHDANOV 377, 391
BABCOCK 20, 21, 56, 57, 67
BAKER 31
BALDET 339
BALDWIN 263
Balmer discontinuity 14
BARABASHOV 242—244, 263, 303, 310, 311, 312, 316, 319, 327
BARNARD 324
BATTEAU 348
BAUM 317
BAUMBACH 153
BEČVÁŘ 153
BEHR 390
BELL 32
BELOPOL'SKII 47, 277, 320
BERNHEIMER 7, 8
BESSEL 343
BEZRUKOVA 64
BIERMANN 59, 67, 105, 110, 367
BILLINGS 76, 89, 109, 159, 160
BIRGE 4
BJERKNES 66
Blackbody 5, 190—191, 206, 207, 295
BLACKWELL 387, 391
BOBROV 321
BOBROVNIKOFF 331
BOGORODSKII 153, 154
Boltzmann's formula 32, 49, 51, 53, 90—95
BOND 324
Bouguer's formula 5, 9
BOWEN 34
BRACEWELL 236
BREDIKHIN 337—339, 347, 391
BREIDO 263
Bremsstrahlung 143, 225—228, 235
BRIGGS 153
Brightness temperature 191—195, 200—201, 210—213, 235
BRILL 11
BROXON 60
BRUNNER 45, 61
BUGOSLAVSKAYA 98, 155, 168
BUMBA 146
BURKE 318
BURNS 20, 69
Butterfly diagram 64—65

CABANNES 34
CAILLATTE 323
CALAMAI 59
CAMICHEL 308
CAMPBELL 69, 320
CANAVAGGIA 11
CARRINGTON 17, 47, 48, 49, 64, 144, 174
CASSINI 238—240, 324
Cassini's division 320
CEPLEHA 379, 380
CHALONGE 11
CHAMBERLAIN 136
CHANDRASEKHAR 264, 293, 327
CHEKIRDA 242
CHEVALIER 39, 45
CHISTYAKOV 248, 263
Chromosphere 68 ff., 111, 163—167; flares. see solar flares; turbulence 88—89, 103—107, 159
CILLIÉ 72, 84, 86, 89, 91, 92, 98, 109
CIMINO 4
CLAAS 32, 33, 34, 35, 77

CLAIRAUT 280
CLEGG 236, 382
CLEMENTI 358
COBLENTZ 317
CODE 318
Color temperature 11
Comets 2, 333—370
COOK 372, 378
CORNU 20
Corona 78, 79, 105—107, 111, 124, 144, 148—168, 178—180, 197—203
condensations, 99, 133, 141, 161—163, 213—215, 223—225, 231—234
Cosmic rays 134, 144—147, 259
Cosmogony 1, 333—335, 369
Counterglow 388, 390
COWLING 59, 67, 147
CSADA 156, 168
CURTIS 69

DABIN 373
DANJON 300, 306
DAVIDSON 69, 70, 89
D'AZAMBUJA 114, 176
DEICH 331
DE JAGER 35
DE LA RUE 111
Dellinger effect 144
DELSEMME 344
DENNING 370
DE OBALDIA 339
DESLANDRES 320
DE VAUCOULEURS 5, 302, 311, 327
DEZSE 172, 181
DOBROVOL'SKII 336, 367, 391
DODSON 136, 147
DOLLFUS 154, 159, 160, 162, 299, 304, 311, 313
DONATI 347
DONN 344
Doppler effect 26—30, 51, 55, 85—86, 89, 99, 126, 129, 130, 138, 159, 162, 274—277, 293, 300, 304, 311, 354
DOUGLAS 318
DUBOV 106, 110, 124, 133
DUFAY 34
DUNÉR 17
DUNHAM 303
DYSON 48, 69

Earth 306—308
Eclipse observations, 183, 196—197, 221, 225, solar 23, 68—70, 79—82, 89ff., 111, 125, 148, 153; lunar 247, 249, 307——308
EDDINGTON 42, 66
EDLÉN 24, 151, 152
Effective temperature 6, 31, 191, 206, 212, 229
Einstein shift 21

ELLERMAN 55, 141, 143
ELLISON 135, 147
ELSASSER 390
EPSTEIN 252
Equivalent widths 25—32, 151—152
EVANS 76, 89, 109
EVERSHED 21, 57—59, 68
Evershed effect 58—59, 61
Excitation temperature 30—31, 84, 89
Extinction 9, 317, 387
EZERSKII 242, 257

FABRICIUS 43
Faculae 41—43, 48—53, 56—57, 63, 102, 139 143, 155, 173—174
FAN 147
Faye's formula 17, 47
FEDORETS 242, 245, 263
FEDOROV 263
FESENKOV 286, 293, 308, 310, 312, 314, 316, 321, 329, 334, 390
FIALKO 391
Filaments, chromospheric 95—98, 102, 112, 114—116, 203—208
FIRSOFF 252
Flash spectrum 68—72, 77
Flocculi, chromospheric 107, 109, 112, 135—137, 173, 176—177, 227—228
FOWLER 68
FRANKLIN 318
FRAUNHOFER 19
Fraunhofer lines 10, 19ff., 37, 52, 58, 60, 68, 77, 83, 99, 136, 143, 149—152, 347, 350, 354 356, 388—389

GAGARIN 307
GALILEO 43, 324, 325
GANSKII 39, 41, 44, 67, 153, 155
GASCOIGNE 5
GAUZIT 34
Geoactivity 144—147, 167, 169, 233
GETMANTSEV 236
GIANUCCI 4
GICLAS 321
GINZBURG 197, 236, 237
GIOVANELLI 93, 109, 146
GLEISSBERG 63
GLOOS 382
GNEVYSHEV 64, 160, 181
GNEVYSHEVA 44, 64, 160, 180, 181
GOLDBERG 31, 77
GOL'DBERG—ROGOZINSKAYA 141
GOPASYUK 141
GOTTSCHALK 31
GRAFF 323
Granulation, photospheric 37—43, 105
GREEN 263
Greenhouse effect 296, 305
GREENHOW 382

GREENSTEIN 356
GROTRIAN 149, 150, 152
Growth curves 28—32, 33, 50—53, 86—87, 102, 129
GRUTUISEN 257
GULDEN 390
GUND 350
GÜNTHER 323
GUREVICH 59, 67
GURZADYAN 236
GUTHNIK 6

HAGEN 93, 109
HALE 54—56, 60, 67, 112
HALL 324
Halley's Comet 330—331, 334, 336, 341, 345 353, 371
HARDY 323
HASER 357, 391
HAWKINS 378
HEDEMAN 136
HELD 28, 87, 109
Helium, discovery of, 111
HERSCHEL, G. 111
HERSCHEL, W. 302, 324
HERZBERG 317, 322, 363
HESS 313, 317
HEWISH 203
HILL 31
HIRAYAMA 328
HOFFMEISTER 370
HOOVER 8
HOUTGAST 35, 60, 76, 109
HOYLE 146
HUBENET 32
HUGGINS 111
HUNAERTS 33, 34, 35, 77, 355, 356
HUNTER 21
HUYGENS 324

Image plane 266
INGHAM 387, 390, 391
INGLIS 109
ISMAILOV 53
IVANIKOV 372, 379, 391
IVANOV-KHOLODNYI 101, 109, 126, 129—153

JACCHIA 377—378
JACOBSON 7
JÄGERMAN 391
JANSSEN 39, 67, 111
JOHNSTON 263
JULIUS 14
Jupiter 277, 294, 296, 297, 314-318, 328, 342

KAIDANOVSKII 296
KALINYAK 159, 253, 257, 263
KALITIN 286
KASHCHEEV 391
KATASEV 391
KEELER 320
KELLOG 327
KEPLER 307
KERIMBEKOV 39
KHABAKOV 255—256, 261, 262, 263
KHAIKIN 236, 296
KHINKULOVA 153, 154
KHOKHLOVA 123, 133, 138, 147
KHVOL'SON 293
KIESS 20
KIMBALL 7
KING 23, 30, 31, 60, 293, 319
KIRCHHOFF 19, 185, 187
KIRKWOOD 320
KOKHAN 263
KONDRAT'EV 353
KONONOVICH 93, 94, 95, 109, 237
KONOPLEVA 361
KORNILOV 51
KOTEL'NIKOV 297
KOTLYAR 67
KOVALEVSKAYA 320
KOZYREV 248, 253, 257, 304, 314
KRAMER 391
KRASOVSKII 4
KRAT, T.V. 45, 49, 50, 52, 67, 108, 109, 162
KRAT, V.A. 18, 35, 66, 67, 76, 89, 93, 104, 108, 109, 134, 136, 137, 147, 148, 151, 152 163, 164, 167, 168, 369
KRINOV 334, 375, 391
KRISTENSEN 16, 76, 109
KRON 7, 353
KUIPER 1, 5, 18, 35, 101, 168, 236, 301, 312, 314, 317, 322, 323, 324, 327
KUN HUANG 158
KUNTZ 153
KUZ'MIN 296

LAGRANGE 369
LAMANSKY 9
Lambert's law 241, 253—254, 284, 331
Landé factor 53
LAPLACE 343, 369
Larmor radius 368
LASSELL 324
LEBEDEV 362
LEBEDINETS 391
LEBEDINSKII 59, 67
LEVIN 343, 345, 361, 362, 372, 391
LILLER 360, 370
Limb darkening 5, 9, 13—16, 60, 192, 312, 317—319
LINDBLAD 16, 76, 109
LINK 292, 302, 308, 327
LIPSKII 263
LOCKYER 68, 111
LODGE 182
LOMMEL 293

Lommel-Seeliger law 253
Lomonosov effect 299, 301, 303
LOVELL 236, 373, 382, 391
LYOT 41, 112, 151, 152, 159, 161, 178, 263, 283, 300, 303
LYUBARSKII 379

MACRIS 41
MAGGINI 316
MAKSUTOV 173
MARKELOVA 263
MARKOV 155, 238, 244, 247–248, 263, 307, 328
MARKOWITZ 345
MAROCHNIK 369, 391
Mars 267, 272, 308–314, 328
MARTIN 197
MARTYNOV 302
MARVIN 7
MASEVICH 314
MAUNDER 47, 64–65, 172, 180
MAXWELL 320, 363
McCRACKEN 373
McCREA 103, 110
McKELLAR 354
McKENNA 147
McLAUGHLIN 313
McMATH 133
MEGGERS 20, 21
MEINEL 147
MELLOTE 218–219
MEL'NIKOV 31, 52, 55, 67, 109
MENZEL 23, 31, 33, 69, 70, 72, 77, 78, 80, 84, 86, 89, 91, 92, 93, 94, 95, 98, 109, 158, 248, 302, 323
Mercury 240, 269, 288, 295, 297, 299–300, 328, 332
Meteors 292, 333, 336, 370–387
Meteorites 242, 245, 258, 329, 334–335
MEYERMANN 4
MICHARD 98, 109
MICHELSON 20
Mie scattering 321, 348, 361
MIKHAILOV 153, 263
MILANKOVIČ 296, 313, 327
MILLMAN 379, 380, 381, 391
MILNE 14, 103, 106, 110, 185, 187
MINKOWSKI 137, 147
MINNAERT 10, 11, 14, 16, 27, 32, 35, 72, 89, 109, 345, 390
MINNETT 215
Minor planets. See Asteroids
MITCHELL 36, 69–70, 77, 78, 79, 109, 126
MITROFANOVA 31, 32, 36
MITROPOL'SKAYA 43, 52, 53, 67
MIYAMOTO 106
MOKHNACH 361
MOLCHANOV 182, 237
MONCINELLI 58
Moon, 238–263 librations 238–240; maria, 240, 245, 254–256, 258–262; other side 259–263; rotation 238–240; surface 252–253, 259; volcanos 256
MOORE 21, 35, 50, 304, 323, 327
MOULTON 390
MULDERS 10, 27, 32, 35, 109
MÜLLER 299, 300, 321
MUSTEL' 1, 36, 43, 52, 106, 107, 108, 110, 138, 147

Neptune 294, 322–323, 369
NEWKIRK 162, 202
NEWTON 47, 67, 114
NICHOLSON 54, 60, 67, 304, 324
NICOLET 6
NIKOL'SKII 101, 109
NIKONOV 153
NIKONOVA 153
NIKULIN 67
NOSKOVA 263
NUNN 47, 67

ÖHMAN 245
OISEAU 348, 359
OL' 64
OLIVIER 370
OORT 343, 369
ÖPIK 305, 372, 374, 375, 380, 391
Ordinary wave 218,–219, 233
ORLOV 48, 341, 391
ORLOVA 285, 307
ORNSTEIN 89

PAGTZOLD 308
PANNEKOEK 72, 86, 91, 109
PANOFSKY 65
PANOVKIN 237
PAN PUH 120
PARENAGO 321
PAULI 362
PAWSEY 236
PECKER 159
PEREPELKIN 109, 171
PERRINE 324
PETROVA 57
PETTENGILL 263
PETTIT 8, 112, 114–123, 133, 150, 244, 248, 249, 252, 304
Pettit's laws 124
PEYTURAUX 10
PHILLIPS 363
PICKERING 321, 324
PIDDINGTON 215, 237, 263
PIERCE 31
PIKEL'NER 163, 166, 167, 168
Planck's law 5, 11, 13, 87, 296

Planets 264ff.; atmosphere 289—294; color index 282—283, 319; disk 265—270; mass 279—281; phase function 281—283; reflectivity 283—285; rotation 272—277, 297—298
PLASKETT 41
Plasma, defined 230—231; oscillations 230——236
Pluto 274, 323, 369
POGSON 111
Polarizing interference filter 112, 136, 160, 163
POLOSKOV 348, 359, 363, 367, 391
POLOZHENTSEVA 242, 243
Poynting-Robertson effect 384, 389
PRAGER 6
PRAVDYUK 109, 147
PROKOF'EV 304
PROKOF'EVA 141, 159, 162
Prominences 42, 61, 68, 111—133, 135—137, 161, 163, 175; classes 116—123; metallic 125—126
PUTILIN 328, 329, 335

RABE 3, 4, 322
Radiation pressure 103, 106, 123, 362—363, 384, 389
Radio emission, lunar 249—250, 252—254; meteoric 381—385; planetary 264, 296—298, 306, 318; solar 13, 94, 95, 101, 144—145, 159, 178—180, 182—237
RADLOVA 244, 315, 319
RAJHL 379, 380
RAMANATHAN 51, 67
RAMSAY 111
RAYET 111
Rayleigh-Jeans formula 187—188, 190, 297
REDMAN 32, 90, 106, 109
REID 147
RENSE 109
Reversing layer 37, 68, 77, 78, 160
Revised Rowland 21—22, 24, 32, 70
RICHARDSON 41, 137, 147, 301, 308
RICHTER 343, 391
RIJVES 343, 361
ROBERTS 76, 89, 98, 109, 159, 161, 162
Rocket measurements 12, 23, 101, 144, 259——263, 373, 378—379, 386—387, 390
ROGGENHAUSEN 180
ROSA 4
ROSS 304
ROSSELAND 103
ROUGIER 244
ROWLAND 25
ROZHDESTVENSKII 30
ROZHKOVSKII 388
RUBASHEV 66
RUBINSTEIN 32
RUSSELL 5, 32, 77, 244, 268, 282, 299, 303, 315, 319, 321, 322
RYLE 236

SAGAN 305, 327
SAHA 69, 103
Saha's formula 32, 49, 51, 53, 84, 90—91, 93, 156
SALOMONOVICH 263, 296
SAMOILOVA-YAKHONTOVA 335
SANDAGE 31
Satellites 324—327
Saturn 277, 294, 297, 318—321
SCHATZEL 322
SCHATZMAN 106
SCHEINER 182
SCHIAPARELLI 313, 347
SCHMIDT 343
SCHÖNBERG 293, 310, 316, 321
SCHRÖTER 302
SCHWABE 61
SCHWARZSCHILD 14, 39, 41, 104—106, 110, 353
SEARES 55
SECCHI 4, 44, 98, 111
Self-absorption 80, 84—89, 102, 126, 129—130, 158
SERVAJEAN 49
SEVERNYI 1, 35, 67, 104, 123, 126, 133, 136, 138, 141, 143, 145, 146, 147, 181
SHAIN 150, 152
SHAPOSHNIKOVA 136
SHARONOV 263—264, 299, 303, 313, 316, 318, 319 327
SHCHEGOLEV 263
SHIFRIN 327
SHKLOVSKII 35, 78, 93, 94, 95, 109, 133, 144, 150, 151, 156, 158, 163, 166—168, 197, 230, 236, 237
SHTAUDE 328
SHUL'MAN 347
SIEDENTOPF 390
SINTON 249, 253, 263, 304, 314
SITNIK 45
SLONIM 62, 67
SMITH 98, 109
SMOLUCHOWSKI 252
SOBOLEV 1, 68, 109, 111, 126, 131, 133, 136, 137, 147, 303, 310, 327
Solar constant 5—8
Solar flares 99, 121, 134—147, 171, 178, 185
Solar parallax 3
Solar rotation 16—18, 20, 25, 41, 46—49, 174
Solar Survey 169—181
SOSNOVA 377
SOUTHWORTH 182
Spicules, chromospheric, 93—100, 104, 155, 163, 204—208
SPÖRER 64, 65, 173, 181
STAIR 263

Stark effect 26, 27, 138–139
STEBBINS 7, 153
Stefan-Boltzmann law 5, 6, 32, 246, 251
STEPANOV 57, 59, 67
STESHENKO 138, 147
STEWART 65
St. JOHN 20, 21, 58—59
STRATONOV 47, 48
STRATTON 69, 70, 89
STRÖMGREN 35
STRUVE, G. 314—315, 320
STRUVE, O 111
SUBBOTIN 327
Sun, chemical composition, 19—25, 77; magnetic field 53—57, 156, 217—218; quiet 186—187, 199, 228
Sunspots 22—23, 37ff., 222—228; classes 46; magnetic fields 54—56, 59—60; migration 64—65
Swan bands 350, 357, 358
SWINGS 344, 353, 356, 357, 391
SYTINSKAYA 263, 289, 310, 311, 327, 372

TANDBERG-HANSEN 126, 132, 133
TEIFEL' 243, 244, 263
TELLER 109
TEN BRUGGENCATE 32, 50—52, 110
TENNANT 111
TERYAEVA 137
THACKERAY 23
THIESSEN 56, 60
THOMAS 90, 104, 110
TIKHOV 150, 168, 267, 307, 310, 314
TITOV 307
Transparency, atmospheric 6—8, 243, 289, 293, 310
TRELLIS, 160, 168
TROITSKII 237, 253, 263, 296
TRUMPLER 273
TSEITLIN 237
TSIMMERMAN 48

UNSÖLD 6, 32–34, 36, 67, 77, 109
Uranus 278, 294, 321–322
UREY 344

VAN DE HULST 32, 101, 149, 154, 156, 159, 167, 168, 202, 203, 233, 387, 390, 391
VAN MAANEN 55
VAN SLUITERS 60
VAN WOERKOM 369
Venus 267, 269, 274, 288, 294, 300—306, 310, 312
VIOLETT 109
VITKEVICH 184, 203
VOIGT 32
VON KLÜBER 50—54, 56, 67
VON SEELIGER 293, 308, 321
VON ZEIPEL 66
VORONTSOV-VEL'YAMINOV 343, 361
VSEKHSVYATSKII 64, 341—345, 391
VYAZANITSYN 3, 19, 31, 68, 76, 82, 91, 98, 109, 126, 129—133, 181

WALDMEIER 4, 8—9, 18, 112, 114, 116, 133, 135, 139, 147, 151, 159, 160—162, 168, 178, 181, 202, 214, 236
WALÉN 66
WALKER 323
WEAVER 21
WEIGERT 344, 345
WESSELINK 251—252
WHIPPLE 343—344, 377—379
WHITFORD 153
WHITNEY 360
Wien's law 11
WILDT 70, 78, 91, 109, 316, 318
WILKINS 260
WILLIAMS 70, 77—78, 109
WILSING 47, 182, 241
WILSON 45—46, 59
WLÉRICK 154
WOLF 61, 171
Wolf number 7—9, 61—65, 113, 171—173
WOLFER 61
WOLLASTON 19
WOLTJER 93, 109
WOOLLEY 5, 78, 90, 101, 109
WRIGHT 31, 32, 312
WURM 353, 360, 362, 363

YAKOVKIN 158
YEH SHIH-HUI 126, 129, 132, 133
Yield function 85—86

ZAKHAROV 390
ZANSTRA 89, 347, 353
ZAVARITSKII 334
Zeeman effect 23, 26, 50, 53, 56, 175, 234
ZEL'TSER 155, 247–248
ZHELEZNYAKOV 236, 237
ZHURAVLEV 52, 55, 67
ZIRIN 126, 132, 133, 140
Zodiacal light 336, 387—390
ZUIKOV 126, 129, 133
ZWAAN 76, 109